(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 April 2005 (14.04.2005)

PCT

(10) International Publication Number
WO 2005/032582 A2

(51) International Patent Classification[7]: A61K 39/09

(21) International Application Number:
PCT/US2004/024868

(22) International Filing Date: 30 July 2004 (30.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/491,822 31 July 2003 (31.07.2003) US
60/541,565 3 February 2004 (03.02.2004) US

(71) Applicant *(for all designated States except US)*: **CHIRON CORPORATION** [US/US]; 4560 Horton Street, Emeryville, CA 94608-2916 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GRANDI, Guido** [IT/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **TELFORD, John** [IT/US]; C/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **BENSI, Giuliano** [IT/US]; C/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(74) Agents: **HALE, Rebecca, M.** et al.; Intellectual Property-R-338, Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMMUNOGENIC COMPOSITIONS FOR STREPTOCOCCUS PYOGENES

(57) Abstract: The invention includes a GAS antigen, GAS 40, which is particularly suitable for use either alone or in combinations with additional GAS antigens, such as GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057.

WO 2005/032582 A2

IMMUNOGENIC COMPOSITIONS FOR *STREPTOCOCCUS PYOGENES*

All documents cited herein are incorporated by reference in their entirety.

CROSS REFERENCE TO RELATED APPLICATIONS, FROM WHICH PRIORITY IS CLAIMED

This application incorporates by reference in their entirety U.S. provisional patent application No. 60/491,822, filed on July 31, 2003, and U.S. provisional patent application No. 60/541,565, filed on February 3, 2004.

FIELD OF THE INVENTION

This invention is in the fields of immunology and vaccinology. In particular, it relates to antigens derived from *Streptococcus pyogenes* and their use in immunisation. All documents cited herein are incorporated by reference in their entirety.

BACKGROUND OF THE INVENTION

Group A streptococcus ("GAS", *S.pyogenes*) is a frequent human pathogen, estimated to be present in between 5-15% of normal individuals without signs of disease. When host defences are compromised, or when the organism is able to exert its virulence, or when it is introduced to vulnerable tissues or hosts, however, an acute infection occurs. Related diseases include puerperal fever, scarlet fever, erysipelas, pharyngitis, impetigo, necrotising fasciitis, myositis and streptococcal toxic shock syndrome.

GAS is a gram positive, non-sporeforming coccus shaped bacteria that typically occurs in chains or in pairs of cells. Although *S.pyogenes* may be treated using antibiotics, a prophylactic vaccine to prevent the onset of disease is desired. Efforts to develop such a vaccine have been ongoing for many decades. While various GAS vaccine approaches have been suggested and some approaches are currently in clinical trials, to date, there are no GAS vaccines available to the public.

It is an object of the invention to provide further and improved compositions for providing immunity against GAS disease and/or infection. The compositions preferably include GAS 40, a GAS virulence factor identified by Applicants, which is particularly suitable for use in vaccines. In addition, the compositions are based on a combination of two or more (*e.g.* three or more) GAS antigens.

SUMMARY OF THE INVENTION

Applicants have discovered a group of thirty GAS antigens that are particularly suitable for immunisation purposes, particularly when used in combinations. In addition, Applicants have identified a GAS antigen (GAS 40) which is particularly immunogenic used either alone or in combinations with additional GAS antigens.

The invention therefore provides an immunogenic composition comprising GAS 40 (including fragments thereof or a polypeptide having sequence identity thereto). A preferred fragment of GAS 40 comprises one or more coiled-coil regions. The invention further includes an immunogenic composition comprising a combination of GAS antigens, said combination consisting of two to ten GAS antigens, wherein said combination includes GAS 40 or a fragment thereof or a polypeptide having sequence identity thereto. Preferably, the combination consists of three, four, five, six, or seven GAS antigens. Still more preferably, the combination consists of three, four, or five GAS antigens.

The invention also provides an immunogenic composition comprising a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of a first antigen group, said first antigen group consisting of: GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057. These antigens are referred to herein as the 'first antigen group'. Preferably, the combination of GAS antigens consists of three, four, five, six, seven, eight, nine, or ten GAS antigens selected from the first antigen group. Preferably, the combination of GAS antigens consists of three, four, or five GAS antigens selected from the first antigen group.

GAS 39, GAS 40, GAS 57, GAS 117, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511 are particularly preferred GAS antigens. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117. Preferably, the combination includes GAS 40.

Representative examples of some of these antigen combinations are discussed below.

The combination of GAS antigens may consist of three GAS antigens selected from the first antigen group. Accordingly, in one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and a third GAS antigen selected from the first antigen group. Preferred combinations include GAS 40, GAS 117 and a third GAS antigen selected from the group consisting of GAS 39, GAS 57, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511.

In another embodiment, the combination of GAS antigens consists of GAS 40 and two additional GAS antigens selected from the first antigen group. Preferred combinations include GAS 40 and two GAS antigens selected from the group consisting of GAS 39, GAS 57, GAS 117, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511. In another embodiment, the combination of GAS antigens consists of GAS 117 and two additional GAS antigens selected from the first antigen group.

The combination of GAS antigens may consist of four GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and two additional GAS antigens selected from the first antigen group. Preferred combinations include GAS 40, GAS 117, and two GAS antigens selected from the group consisting of GAS 39, GAS 57, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511.

In another embodiment, the combination of GAS antigens consists of GAS 40 and three additional GAS antigens selected from the first antigen group. Preferred combinations include GAS 40 and three additional GAS antigens selected from the group consisting of GAS 39, GAS 57, GAS 117, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511. In one embodiment, the combination of GAS antigens consists of GAS 117 and three additional antigens selected from the first antigen group.

The combination of GAS antigens may consist of five GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and three additional GAS antigens selected from the first antigen group. Preferred combinations include GAS 40, GAS 117 and three additional GAS antigens selected from the group consisting of GAS 39, GAS 57, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511.

In another embodiment, the combination of GAS antigens consists of GAS 40 and four additional GAS antigens selected from the first antigen group. Preferred combinations include GAS 40 and four additional GAS antigens selected from the group consisting of GAS 39, GAS 57, GAS 117, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511. In one embodiment, the combination of GAS antigens consists of GAS 117 and four additional GAS antigens selected from the first antigen group.

The combination of GAS antigens may consist of eight GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and six additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40 and seven additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 117 and seven additional GAS antigens selected from the first antigen group.

The combination of GAS antigens may consist of ten GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and eight additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40 and nine additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 117 and nine additional GAS antigens selected from the first antigen group.

BRIEF DESCRIPTION OF THE FIGURES

FIGURE 1 identifies a leader peptide sequence, two coiled-coil sequences, a leucine zipper sequence and a transmembrane sequence within a GAS 40 amino acid sequence.

FIGURE 2 depicts a schematic of GAS 40 identifying a leader peptide sequence, two coiled-coil sequences, a leucine zipper sequence and a transmembrane sequence, as well as coiled-coil regions of GAS 40 which have low level homology with other Streptococcal proteins of known or predicted function.

FIGURE 3 includes the BLAST alignment analysis of identified coiled-coil regions of GAS 40 with other Streptococcus bacteria.

FIGURE 4 provides predicted secondary structure for an amino acid sequence of GAS 40.

FIGURE 5 schematically depicts the location of GAS 40 within the GAS genome. It also includes comparison schematic depicting a GAS mutant with GAS 40 deleted. Further details on these schematics demonstrate the likelihood that GAS 40 was acquired by horizontal transfer through a transposon factor.

FIGURE 6 provides comparison FACS analysis depicting the surface exposure of GAS 40 in a wild type strain (and no surface exposure in the GAS 40 deletion mutant).

FIGURE 7 presents opsonophagocytosis data for GAS 40 (in various expression constructs).

FIGURE 8 presents immunization and challenge data for several GAS antigens of the invention.

DETAILED DESCRIPTION OF THE INVENTION

As discussed above, the invention provides compositions comprising a combination of GAS antigens, wherein the combinations can be selected from groups of antigens which Applicants have identified as being particularly suitable for immunization purposes, particularly when used in combination. In particular, the invention includes compositions comprising GAS 40.

GAS 40 and the other GAS antigens of the first antigen group are described in more detail below. Genomic sequences of at least three GAS strains are publicly available. The genomic sequence of an M1 GAS strain is reported at Ferretti et al, PNAS (2001) 98(8):4658 – 4663. The genomic sequence of an M3 GAS strain is reported at Beres et al., PNAS (2002) 99(15):10078 – 10083. The genomic sequence of an M18 GAS strain is reported at Smooet et al., PNAS (2002) 99(7):4668 – 4673. Preferably, the GAS antigens of the invention comprise polynucleotide or amino acid sequence of an M1, M3 or M18 GAS strains. More preferably, the GAS antigens of the invention comprise a polynucleotide or amino acid sequence of an M1 strain.

As there will be variance among the identified GAS antigens between GAS M types and GAS strain isolates, references to the GAS amino acid or polynucleotide sequences of the invention preferably include amino acid or polynucleotide sequences having sequence identity thereto. Preferred amino acid or polynucleotide sequences have 50% or more sequence identity (*e.g.*, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more). Similarly, references to the GAS amino acid or polynucleotide sequences of the invention preferably include fragments of those sequences, (*i.e.*, fragments which retain or encode for the immunological properties of the GAS antigen). Preferred amino acid fragments include at least *n* consecutive amino acids, wherein *n* is 7 or more (*e.g.*, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). Preferred polynucleotide fragments include at least n consecutive polynucleotides, wherein n is 12 or more (*e.g.*, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). In one embodiment, the amino acid or polynucleotide fragments of the invention are not identical to amino acid or polynucleotide sequences from other (non-GAS) bacteria (e.g., the fragments are not identical to sequences in other *Streptococcus* bacteria).

(1) GAS 40

GAS 40 corresponds to M1 GenBank accession numbers GI:13621545 and GI:15674449, to M3 GenBank accession number GI: 21909733, to M18 GenBank accession number GI:19745402, and is also referred to as 'Spy0269' (M1), 'SpyM3_0197' (M3), 'SpyM18_0256' (M18) and 'prgA'. GAS 40 has also been identified as a putative surface exclusion protein. Amino acid and polynucleotide sequences of GAS 40 from an M1 strain are set forth below and in the sequence listing as SEQ ID NOS: 1 and 2.

SEQ ID NO: 1

MDLEQTKPNQVKQKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVEKTLSQQKAE
LTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATETELHNAQADQHSKETALS
EQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEEL
AAQKAALAEKEAELSRLKSSAPSTQDSIVGNNTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQL
NQYQDIPADRNRFVDPDNLTPEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQP
GVSGHYGVGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLRVD
KHNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGKVSSLENRLSAI
HQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLASLKAALHQTE
ALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEH
QLTLLKTLANEKEYRHLDEDIATVPDLQVAPPLTGVKPLSYSKIDTTPLVQEMVKETKQLLEASARLAAENTSLVAEALV
GQTSEMVASNAIVSKITSSITQPSSKTSYGSGSSTTSNLISDVDESTQRALKAGVVMLAAVGLTGFRFRKESK

SEQ ID NO: 2

ATGGACTTAGAACAAACGAAGCCAAACCAAGTTAAGCAGAAAATTGCTTTAACCTCAACAATTGCTTTATTGAGTGCCAG
TGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATACTCACGACGATAGTT
TACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAACTCTCAGTCAACAAAAAGCAGAA

CTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAATGAACAAAA
AGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAAC
ATCAAAGAGAGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCA
GAACAAAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATAT
TGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCAT
TAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAAGAGTTG
GCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAGCTCCGTCTACTCAAGATAG
CATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAACTTAAAAAATTAGAAGCTAGTGGTTATA
TTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTA
AATCAATACCAAGATATTCCAGCAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCT
AGCGCAGTTTGCAGCTCACATGATTAATAGTGTAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCAC
AAGAATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCA
GGGGTATCAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATTCG
AAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACGTGGTATTTATG
ACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTATTAACTTTTTACGTGTAGAT
AAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTT
TCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCC
AAAGAGTAGGCACTGTATCTGATACTATTGCAGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATT
CATCAAGAAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTC
AGACAGCTTAAATCTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAA
AACAAGCACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAA
GCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCTAAGGGACTTTAA
ATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTTGGCTAAAACTACCTCATCTT
TGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGTCTAGAAGCTACTATTGCTACCACAGAACAC
CAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTT
GCAAGTAGCTCCACCTCTTACGGGCGTAAAACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGG
TTAAAGAAACGAAACAACTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTT
GGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGAC
ATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAAGAGCTCTTAAAGCAG
GAGTCGTCATGTTGGCAGCTGTCGGCCTCACAGGATTTAGGTTCCGTAAGGAATCTAAGTGA

Preferred GAS 40 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 1; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 1, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). These GAS 40 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 1. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 1. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 1.

For example, in one embodiment, the underlined amino acid sequence at the N-terminus (leader sequence) of SEQ ID NO: 1 is removed. (The amino acid and polynucleotide sequences for this N terminal leader sequence are listed in the sequence listing as SEQ ID NOS: 3 and 4. The amino acid and polynucleotide sequences for the remaining GAS 40 fragment are listed in the sequence listing as SEQ ID NOS: 5 and 6.)

As another example, in one embodiment, the underlined amino acid sequence at the C-terminus (transmembrane region) of SEQ ID NO: 1 is removed. (The amino acid and polynucleotide sequences for this transmembrane region are listed in the sequence listing as SEQ ID NOS: 7 and 8. The amino acid and polynucleotide sequences for the remaining GAS 40 fragment are listed in the sequence listing as SEQ ID NOS: 9 and 10).

Other fragments may omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

Further illustration of domains within GAS 40 is shown in FIGURES 1 and 2. As shown in these figures, an amino acid sequence for GAS 40 (SEQ ID NO: 1) contains a leader peptide sequence within amino acids 1 – 26 (for example SEQ ID NO: 3), a first coiled-coil region within amino acids 58 – 261 (SEQ ID NO: 12), a second coiled coil region generally within amino acids 556 – 733 (SEQ ID NO: 13), a leucine zipper region within amino acids 673 – 701 (SEQ ID NO: 14) and a transmembrane region within amino acids 855 – 866 (SEQ ID NO: 11). Figure 1 depicts these regions within an amino acid sequence for GAS 40, while Figure 2 depicts these regions schematically along the length of the GAS 40 protein.

The coiled-coil regions identified within GAS 40 are likely to form alpha helical coils. These structures are frequently involved in oligomerization interactions, for example between different regions of the protein or between regions of two separate proteins. The leucine zipper motif within the second coiled-coil region contains a series of leucine (or isoleucine) amino acid residues, spaced in such a way as to allow the protein to form a specialized oligomerization interaction between two alpha helices. In a leucine zipper motif, preferably, there are six amino acid residues interspaced between the repeating leucine residues. In a leucine zipper oligomeric structure, the alpha helices are thought to be held together by hydrophobic interactions between leucine residues, which are located on one side of each helix. Leucine zipper motifs are frequently involved in dimerization interactions. The location of the leucine zipper motif within the coiled-coil region further indicates the likelihood that this region of the GAS 40 protein is involved in an oligomerization interaction.

FIGURE 2 also illustrates that there is low level homology between some of the identified regions of GAS 40 and other Streptococcal proteins with known or predicted two dimensional structures or surface localization. Such low level homology may indicate a similar secondary structures or even function. For example, amino acids 33 to 324 of GAS 40, including the first coiled-coil region, has approximately 22% sequence identity to a region (amino acids 112 to 392) of a protein from *Streptococcus gordonii* called streptococcal surface protein A ("SpA") precursor (Genbank reference GI 25990270, SEQ ID NO: 15). This protein is thought to be a surface protein adhesion, involved in the adhesion of that Streptococcus with mammalian host cell membranes. The *S. gordonii* SpA is a member of streptococcal antigen I/II family of protein adhesions and recognizes salivary agglutinin glycoprotein (gp-340) and type I collagen. Amino acids 33 to 258 of GAS 40 also show low level sequence identify (23%) with another S. gordonii protein, Streptococcal surface protein B precursor (Genbank reference GI 25055226, SEQ ID NO: 16).

A similar region of GAS 40 which also overlaps with the first coiled-coil region (amino acids 43 – 238) demonstrates about 23% sequence identity to a region (amino acids 43 — 238) of a protein from *Streptococcus pneumoniae* called surface protein pspA precursor (Genbank reference GI 282335, SEQ ID NO: 17). The aminoterminal domain of pspA is thought to be essential for full pneumococcal virulence, and monoclonal antibodies raised against it protect mice against pneumococcal infections. The pspA domain has a monomeric form with an axial shape ratio of

approximately 1:12, typical of fibrous proteins. Sequence analyses indicates an alpha-helical coiled-coil structure for this monomeric molecule with only few loop-type breaks in helicity.

The second coiled-coil region of GAS 40 has about 46% sequence identity to a region (amino acids 509 – 717) of a protein from *Streptococcus equi* called immunoreactive protein Se89.9 (Genbank reference GI 2330384, SEQ ID NO: 18) (the full length sequence for Se89.9 is also available at http://pedant.gsf.de). This *Streptococcus equi* protein is predicted to be surface exposed. BLAST alignment of each of these Streptococcal sequences with GAS 40 is presented in Figure 3.

Further illustration of the two dimensional structure of GAS 40 is shown in Figure 4. First, Figure 4(a) presents predicted secondary structure analysis aligned against the amino acid sequence for GAS 40. The predicted alpha helical regions in Figure 4 generally correspond to the previously noted coiled-coil regions. In Figure 4(b), PairCoil prediction is used to predict the location of putative coiled-coils. Here, two coil regions are identified, generally corresponding to the first and second coiled coil regions. Figure 4(c) highlights the leucine zipper region and illustrates the regularly repeating leucine (or isoleucine) amino acid residues which are likely to participate in the leucine zipper.

Accordingly, the first coiled-coil region of GAS 40 comprises an amino acid sequence of at least ten (*e.g.*, at least 10, 13, 15, 18, 20, 25, 30, 35, 40, 50, 70, 90, 100 or more) consecutive amino acid residues, selected from the N-terminal half of a full length GAS 40 sequence, and predicted to form an alpha-helical complex based on the functional characteristics of the amino acid residues in the sequence. SEQ ID NO: 12 is a preferred first coiled-coil region of GAS 40.

The second coiled-coil region of GAS 40 comprises an amino acid sequence of at least ten (*e.g.*, at least 10, 13, 15, 18, 20, 25, 30, 35, 40, 50, 70, 90, 100 or more) consecutive amino acid residues, selected from the C-terminal half of a full length GAS 40 sequence, and predicted to form an alpha-helical complex based on the functional characteristics of the amino acid residues in the sequence. The second coiled-coil region preferably includes a leucine zipper motif. SEQ ID NO: 13 is a preferred second coiled–coil region of GAS 40.

The coiled-coil regions of GAS 40 are likely to be involved in the formation of oligomers such as dimers or trimers. Such oligomers could be homomers (containing two or more GAS 40 proteins oligomerized together) or heteromers (containing one or more additional GAS proteins oligomerized with GAS 40). Alternatively, the first and second coiled-coil regions may be interacting together within the GAS 40 protein to form oligomeric reactions between the first and second coiled-coil regions.

Accordingly, in one embodiment, the compositions of the invention include a GAS 40 antigen in the form of an oligomer. The oligomer may comprise two more GAS 40 antigens or fragments thereof, or it may comprise GAS 40 or a fragment thereof oligomerized to a second GAS antigen. Preferred GAS 40 fragments comprise an amino acid sequence selected from the group consisting of the first coiled-coil region

and the second coiled-coil region. Such preferred GAS 40 fragments may be used alone or in the combinations of the invention.

The GAS polynucleotides and amino acid sequences of the invention may be manipulated to facilitate or optimise recombinant expression. For example, the N-terminal leader sequence may be replaced with a sequence encoding for a tag protein such as polyhistidine ("HIS") or glutathione S-transferase ("GST"). Such tag proteins may be used to facilitate purification, detection and stability of the expressed protein. Variations of such modifications for GAS 40 are discussed below. Such modifications can be applied to any of the GAS proteins of the invention.

An example of a GAS 40 sequence with both a GST and a HIS tag is denoted herein as "GST 40 HIS". This construct includes a GAS 40 sequence where the leader sequence is removed, a GST tag coding sequence is added to the N-terminus, and a HIS tag coding sequence is added to the C-terminus (using, for example, a pGEXNNH vector with NdeI and NotI restriction sites). Polynucleotide and amino acid sequences for the fused region of the GST tag, the GAS 40 sequence and the C-terminus HIS tag of GST 40 HIS are shown in SEQ ID NOS: 19 and 20.

Alternatively, a single tag sequence may be used. An example of a GAS 40 sequence with just a HIS tag is denoted as "40a-HIS". This construct includes a GAS 40 sequence where the N-terminus leader sequence and the C-terminus containing the transmembrane sequence is removed. In this construct, the HIS tag sequence is added to the C-terminus (using for example, a cloning vector such as pET21b+ (Novagen) at the NdeI and NotI restriction sites). Polynucleotide and amino acid sequences for 40a-HIS are shown in SEQ ID NOS. 21 and 22.

In addition to the addition of purification tags, recombinant expression may also be facilitated by optimising coding sequences to those more abundant or accessible to the recombinant host. For example, the polynucleotide sequence AGA encodes an arginine amino acid residue. Arginine may also be encoded by the polynucleotide sequence CTG. This CTG codon is preferred by the translational enzymes in *E. coli*. In the 40a-HIS polynucleotide sequence SEQ ID NO 21, a C-terminus CTG coding for arginine has been replaced with CGT.

The following codons are generally underrepresented in E.coli: AGA, AGG and CGA. When these codons occur in a GAS polynucleotide sequence, they may be replaced with one of the other two optional codons encoding for the same amino acid residue.

A total of three ATG codons are optimised to CTG in the "40a-RR-HIS" construct, SEQ ID NOS 23 and 24. SEQ ID NO 23 is also shown below, with the optimised codons underlined. (other than the additional codon optimisation, 40a-RR-HIS is identical to 40a-HIS.)

SEQ ID N: 23

ATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATACTCACGACG
ATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAACTCTCAGTCAACAAAA
AGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAAT
GAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAG
GAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGAC
TGCATTGTCAGAACAAAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCT
GAACAAAATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATA
ATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATT

GACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAGCTCCG
TCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAACTTAAAAAATTAG
AAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAG
TCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCA
GAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAATAGTGTA**cGtcGt**CAATTAGGTCTACCACCAGTTA
CTGTTACAGCAGGATCACAAGAATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATC
ATTTGTCTACGGACAGCCAGGGGTATCAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCC
GGAGCGTCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATG
GTATTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGC
TATTAACTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTAGGATCT
TTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCTATAAAAGCCG
TTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAGCGATCAAAGGAAAAGTAAGCTC
ATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGT
AAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGA
GAACAGAATTACTAGCAGCTAAAGCAAAACAAGCACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATC
GTTGAAAGCCGCACTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAA
GCTCATTTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTA
AGCAAGATTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAG
TCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAATATCGCCAC
TTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAAACCGCTATCATATAGTA
AGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACTATTAGAAGCTTCAGCAAGATTAGCTGC
TGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAA
ATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTG
ATGTTGATGAAAGTACTCAA**cGtGCGGCCGCACTCGAGCACCACCACCACCACCACCAC**

Codon optimisation can also be used without a purification tag. Construct "40a-RR-Nat", SEQ ID NOS: 25 and 26, provides such an example. This construct comprises GAS 40 without the N-terminus leader sequence and the C-terminus transmembrane sequence, with three codon optimisations (and does not include a HIS tag sequence).

Different cloning vectors can be used to optimise expression in different host cells or under different culture conditions. The above discussed constructs used pET21b+ (Novagen) vector which includes an IPTG inducible promoter. As an alternative, an *E.coli/B.subtilis* expression shuttle vector such as pSM214gNH may be used. This vector uses a constitutive promoter instead of an IPTG inducible promoter. An example of a GAS 40 construct using this vector is denoted as "HIS-40a-NH", SEQ ID NOS 27 and 28. In this construct, both the N-terminus leader sequence and the C-terminus transmembrane sequence are removed, and a HIS tag is added to the N-terminus. Additional N-terminus amino acids are introduced with the cloning. In addition, two nucleotide changes which most likely occurred during PCR are indicated – neither of these changes results in amino acid changes.

As another alternative, the pSM214gCH shuttle vector may be used. An example of a GAS 40 construct using this vector is denoted as "HIS-40a-CH", SEQ ID NOS: 29 and 30. In this construct, the N-terminus leader sequence and the C-terminus transmembrane sequence are removed and the HIS tag is placed at the C-terminus. Two additional amino acids are also introduced at the amino terminus. Three nucleotide changes introduced with the cloning are shown in the DNA sequence, with a resulting amino acid change indicated in the protein sequence (from amino acid F to S).

Codon optimisation can also be used with these alternative cloning vectors. GAS 40 construct "HIS- 40a-RR-NH" comprises the "HIS-40a-NH" construct with three codon optimisations. HIS-40a-RR-NH is set forth in the sequence listing as SEQ ID NOS: 31 and 32.

Accordingly, the GAS antigens used in the invention may be produced recombinantly using expression constructs which facilitate their recombinant production. Preferred sequence modifications to facilitate expression may be selected from the group consisting of (1) the addition of a purification tag sequence and (2) codon optimisation.

As discussed above, Applicants have identified GAS 40 as being particularly suitable for use in immunogenic compositions, either alone or in combinations. The use of GAS 40 as a particularly effective GAS antigen is supported by its association with virulence, its surface localization, its effectiveness in bacterial opsonophagocytosis assays and in immunization challenge experiments. In addition, the potential horizonatal acquisition of this virulence factor indicates that this antigen may be specific to GAS (relative to other Streptococcal bacteria). Further support for the antigenic properties of GAS 40 also includes the identification of coiled-coil regions within the GAS 40 two dimensional structure, and the low level homology of these regions with surface proteins of other Streptococcal bacteria, including some adhesion proteins.

Applicants' analysis of the location of GAS 40 within the *Streptococcal pyogenes* genome indicates that this virulence factor was likely acquired by GAS during evolution as a result of a horizontal gene transfer. Figure 5A depicts GAS 40 within the GAS genome. It is preceded on the 5' end by a sequence designated "purine operon repressor" or "purR". It is followed on the 3' end by two sequences encoding ribosomal proteins designated "ribosomal protein S12", or "rpsL" and "ribosomal protein S7" or "rpsG". (Amino acid and polynucleotide sequences for these flanking genes are publicly available on GenBank. (PurR sequences can be found for example under Genbank reference GI:15674250. RpsL sequences can be found for example under Genbank reference GI:15674250. RpsG sequences can be found for example under Genbank reference GI:15674250. Notably, there are two putative promoter sequences designated at the beginning of the rpsL sequence. Figure 5B depicts a GAS mutant where a large portion of GAS 40 is deleted. The only portionof the GAS 40 sequence remaining corresponds to polynucleotides 1 – 97 of SEQ ID NO: 2. The deletion included one of the rpsL promoters, leaving the second, P*, intact. (The horizontal arrows underlining the schematic indicate the deleted region.)

Figure 5C provides additional detail on the wildtype GAS sequence. Here, direct repeat sequences, designated "DR", are shown flanking the 5' and 3' ends of GAS 40. (The corresponding sequences in the GAS 40 deletion mutant are identified in Figure 5D). These direct repeat sequences are approximately 8 basepairs. One example of such a basepair direct repeat comprises SEQ ID NO: 136. Such sequence motifs within a bacterial genome frequently indicate a horizontal gene transfer. *In vivo* infection experiments show that the GAS 40 deletion mutant is several logs less virulent than the wild type strain. (Details of this experiment are provided in Example 2).

The combination of the presence of the flanking direct repeat sequences and the virulence associated with GAS 40 strongly suggests that the GAS 40 sequence was horizontally acquired by *Streptococcus pyogenes* during evolution. Notably, while related purR and rpsL are present in related

Streptococcal bacteria *Streptococcus agalactiae* and *Streptococcus mutants*, neither of these bacteria are known to have a GAS 40 homologue. (Figure 5E schematically depicts the location of purR, rpsL, and rpsG homologues within *S. agalactiae* (Group B Streptococcus) and shows the percent homology of the GBS homologues with the GAS counterparts. Notably, GBS genomes generally only possess one of the direct repeat sequences – and do not contain a pair of the direct repeat sequences flanking the GAS 40 sequence.)

The surface location of GAS 40 is illustrated by the FACS diagram presented in Figure 6. (Discussion of protocols relating to FACS analysis is presented in Example 1). Figure 6 includes FACS diagrams for both the wild type GAS (designated DSM 2071, an M23 type of GAS) and the deletion mutant (designated DSM 2071Δ40). The absorbance shift for the wild type strain indicates that GAS 40 is recognized on the surface of the bacteria by anti-GAS 40 antibodies (and that it is not recognized on the surface of the deletion mutant).

The surface exposure of GAS 40 is further demonstrated by a bacterial opsonophagocytosis assay illustrated in Figure 7 and in Example 3. In this assay, GAS strains are incubated with preimmune and immune sera, polymorphonucleates and complement. (The immune sera is generated by mouse immunization with the indicated GAS protein.) Phagocytosis or growth of the bacteria are measured logarithmically. Positive histogram bars represent phagocytosis (or bacterial death). Negative histogram bars represent bacterial growth. As shown in Figure 7, immune sera generated by each of the GAS40 expressed proteins resulted in a reduction of bacteria (positive histogram bars).

Immunization challenge studies with GAS 40 are discussed in detail in Example 4. As shown in this example, GAS 40, as produced using various constructs, provides substantial protection in adult mice. Notably, most GAS40 constructs provide almost as much protection as GAS M protein. (GAS M protein is used for comparison as it is known to be highly immunogenic. However, M protein is generally not regarded as a suitable GAS vaccine candidate as it varies widely among GAS strains and has epitopes with potential cross-reactivity with human tissues.) In addition, an N-terminus fragment of GAS 40 also provided significant protection in this model. The N-terminus fragment comprises about 292 amino acids from the N-terminus of GAS 40 overlaps with the first coiled-coil region. "40N-HIS" (SEQ ID NOS. 33 and 34) is an example of this GAS 40 fragment which comprises the coiled-coil region of GAS 40 and a C-terminus HIS tag.

(2) GAS 117

GAS 117 corresponds to M1 GenBank accession numbers GI:13621679 and GI:15674571, to M3 GenBank accession number GI:21909852, to M18 GenBank accession number GI: 19745578, and is also referred to as 'Spy0448' (M1), 'SpyM3_0316' (M3), and 'SpyM18_0491' (M18). Examples of amino acid and polynucleotide sequences of GAS 117 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 35 and 36.

Preferred GAS 117 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%,

98%, 99%, 99.5% or more) to SEQ ID NO: 35; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 35, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). These GAS 117 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 35. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 1. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 35. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 35 (shown below) is removed. (SEQ ID NO: 37 comprises the removed N-terminal amino acid sequence. SEQ ID NO: 38 comprises a fragment of GAS 117 without the N-terminal amino acid sequence). Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 35

MTLKKHYYLLSLLALVTVGAAFNTSQSVSAQVYSNEGYHQHLTDEKSHLQYSKDNAQLQLRNILDGYQNDLGRHYSSYYY
YNLRTVMGLSSEQDIEKHYEELKNKLHDMYNHY

(3) GAS 130

GAS 130 corresponds to M1 GenBank accession numbers GI:13621794 and GI:15674677, to M3 GenBank accession number GI: 21909954, to M18 GenBank accession number GI: 19745704, and is also referred to as 'Spy0591' (M1), 'SpyM3_0418' (M3), and 'SpyM18_0660' (M18). GAS 130 has potentially been identified as a putative protease. Examples of amino acid and polynucleotide sequences of GAS 130 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 39 and 40.

Preferred GAS 130 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 39; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 39, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, or more). These GAS 130 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 39. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 39. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 39. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(4) GAS 277

GAS 277 corresponds to M1 GenBank accession numbers GI:13622962 and GI:15675742, to M3 GenBank accession number GI: 21911206, to M18 GenBank accession number GI: 19746852, and is also referred to as 'Spy1939' (M1), 'SpyM3_1670' (M3), and 'SpyM18_2006' (M18). Amino acid and polynucleotide sequences of GAS 277 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 41 and 42.

Preferred GAS 277 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 41; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 41, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These GAS 277 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 41. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 41. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 41. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 41 (shown below) is removed. (SEQ ID NO: 43 comprises the underlined N-terminal amino acid. SEQ ID NO: 44 comprises a fragment of GAS 277 with the N-terminal amino acid sequence removed). Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 41

MTTMQKTISLLSLALLIGLLGTSGKAISVYAQDQHTDNVIAESTISQVSVEASMRGTEPYIDATVTTDQPVRQPTQATIT
LKDASDNTINSWVYTMAAQQRRFTAWFDLTGQKSGDYHVTVTVHTQEKAVTGQSGTVHFDQNKARKTPTNMQQKDTSKAM
TNSVDVDTKAQTNQSANQEIDSTSNPFRSATNHRSTSLKRSTKNEKLTPTASNSQKNGSNKTKMLVDKEEVKPTSKRGFP
WVLLGLVVSLAAGLFIAIQKVSRRK

(5) GAS 236

GAS 236 corresponds to M1 GenBank accession numbers GI:13622264 and GI:15675106, M3 GenBank accession number GI: 21910321, and to M18 GenBank accession number GI: 19746075, and is also referred to as 'Spy1126' (M1), 'SpyM3_0785' (M3), and 'SpyM18_1087' (M18). Amino acid and polynucleotide sequences of GAS 236 from an M1 strain are set forth in the sequence listing as SEQ ID NOS: 45 and 46.

Preferred GAS 236 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 45; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 45, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS 236 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 45. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 45. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 45. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 45 (shown below) is removed. (SEQ ID NO: 47 comprises the N-terminus amino acid sequence. SEQ ID NO: 48 comprises a fragment of GAS 236 with the N-terminus sequence removed). Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 45

MTQMNYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGMLLSAFHMYEKELDKVRFVGIHT
GHLGFYTDYRDFEVDKLIDNLRKDKGEQISYPILKVAITLDDGRVVKARALNEATVKRIEKTMVADVIINHVKFESFRGD

GISVSTPTGSTAYNKSLGGAVLHPTIEALQLTEISSLNNRVFRTLGSSIIIPKKDKIELVPKRLGIYTISIDNKTYQLKN
VTKVEYFIDDEKIHFVSSPSHTSFWERVKDAFIGEIDS

(6) GAS 389

GAS 389 corresponds to M1 GenBank accession numbers GI:13622996 and GI:15675772, to M3 GenBank accession number GI: 21911237, to M18 GenBank accession number GI: 19746884, and is also referred to as 'Spy1981' (M1), 'SpyM3_1701' (M3), 'SpyM18_2045' (M18) and 'relA'. GAS 389 has also been identified as a (p)ppGpp synthetase. Amino acid and polynucleotide sequences of GAS 389 from an M1 strain are set forth in the sequence listing as SEQ ID NOS: 49 and 50.

Preferred GAS 389 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 49; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 49, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). These GAS 389 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 49. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 49. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 49. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(7) GAS 504

GAS 504 corresponds to M1 GenBank accession numbers GI:13622806 and GI:15675600, to M3 GenBank accession number GI: 21911061, to M18 GenBank accession number GI: 19746708, and is also referred to as 'Spy1751' (M1), 'SpyM3_1525', 'SpyM18_1823' (M18) and 'fabK'. GAS 504 has also been identified as a putative trans-2-enoyl-ACP reductase II. Amino acid and polynucleotide sequences of GAS 504 of an M1 strain are set forth below and in the sequence listing as SEQ ID NOS: 51 and 52.

Preferred GAS 504 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 51; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 51, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS 504 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 51. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 51. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 51. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(8) GAS 509

GAS 509 corresponds to M1 GenBank accession numbers GI:13622692 and GI:15675496, to M3 GenBank accession number GI: 21910899, to M18 GenBank accession number GI: 19746544, and is also referred to as 'Spy1618' (M1), 'SpyM3_1363' (M3), 'SpyM18_1627' (M18) and 'cysM'. GAS 509 has

also been identified as a putative O-acetylserine lyase. Amino acid and polynucleotide sequences of GAS 509 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 53 and 54.

Preferred GAS 509 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 53; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 53, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These GAS 509 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 53. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 53. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 53. For example, in one embodiment, the underlined amino acid sequence at the C-terminus of SEQ ID NO: 53 (shown below) is removed. (SEQ ID NO: 55 comprises the C-terminus amino acid sequence. SEQ ID NO: 56 comprises a fragment of GAS 509 with the C-terminus sequence removed). Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 53

MTKIYKTITELVGQTPIIKLNRLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEGLISPGDVIIE
PTSGNTGIGLAWVGAAKGYRVIIVMPETMSLERRQIIQAYGAELVLTPGAEGMKGAIAKAETLAIELGAW
MPMQFNNPANPSIHEKTTAQEILEAFKEISLDAFVSGVGTGGTLSGVSHVLKKANPETVIYAVEAEESAV
LSGQEPGPHKIQGISAGFIPNTLDTKAYDQIIRVKSKDALETARLTGAKEGFLVGISSGAALYAAIEVAK
QLGKGKHVLTILPDNGERYLSTELYDVPVIKTK

(9) GAS 366

GAS 366 corresponds to M1 GenBank accession numbers GI:13622612, GI:15675424 and GI:30315979, to M3 GenBank accession number GI: 21910712, to M18 GenBank accession number GI: 19746474, and is also referred to as 'Spy1525' (M1), 'SpyM3_1176' (M3), 'SpyM18_1542' (M18) and 'murD'. GAS 366 has also been identified as a UDP-N-acetylemuramoylalanine-D-glutamate ligase or a D-glutamic acid adding enzyme. Amino acid and polynucleotide sequences of GAS 366 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 57 and 58.

Preferred GAS 366 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 57; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 57, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS 366 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 57. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 57. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 57. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 57 (shown below) is removed. (SEQ ID NO: 59 comprises the N-terminus leader sequence. SEQ ID NO: 60 comprises a fragment of GAS 366 where the N-terminus sequence is removed). Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 57

MKVISNFQNKKILILGLAKSGEAAAKLLTKLGALVTVNDSKPFDQNPAAQALLEEGIKVICGSHPVELLDENFEYMVKNP
GIPYDNPMVKRALAKEIPILTEVELAYFVSEAPIIGITGSNGKTTTTTMIADVLNAGGQSALLSGNIGYPASKVVQKAIA

GDTLVMELSSFQLVGVNAFRPHIAVITNLMPTHLDYHGSFEDYVAAKWMIQAQMTESDYLILNANQEISATLAKTTKATV
IPFSTQKVVDGAYLKDGILYFKEQAIIAATDLGVPGSHNIENALATIAVAKLSGIADDIIAQCLSHFGGVKHRLQRVGQI
KDITFYNDSKSTNILATQKALSGFDNSRLILIAGGLDRGNEFDDLVPDLLGLKQMIILGESAERMKRAANKAEVSYLEAR
NVAEATELAFKLAQTGDTILLSPANASWDMYPNFEVRGDEFLATFDCLRGDA

(10) GAS 159

AS 159 corresponds to M1 GenBank accession numbers GI:13622244 and GI:15675088, to M3 GenBank accession number GI: 21910303, to M18 GenBank accession number GI: 19746056, and is also referred to as 'Spy1105' (M1), 'SpyM3_0767' (M3), 'SpyM18_1067' (M18) and 'potD'. GAS 159 has also been identified as a putative spermidine/putrescine ABC transporter (a periplasmic transport protein). Amino acid and polynucleotide sequences of GAS 159 of an M1 strain are set forth below and in the sequence listing as SEQ ID NOS: 61 and 62.

SEQ ID NO: 61

MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSNEAMYTKIKQ
GGTTYDIAVPSDYTIDKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPQNDYSLPYFWGTVGIVYNDQLVDKAPMHWED
LWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNVKAIVADEMKGYMIQGDAAIGITFSGEA
SEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFLNFINRPENAAQNAAYIGYATPNKKAKALLPDEIKNDPA
FYPTDDIIKKLEVYDNLGSRWLGIYNDLYLQFKMYRK

SEQ ID NO: 62

ATGCGTAAACTTTATTCCTTTCTAGCAGGAGTTTTGGGTGTTATTGTTATTTTAACAAGTCTTTCTTTCATCTTGCAGAA
AAAATCGGGTTCTGGTAGTCAATCGGATAAATTAGTTATTTATAACTGGGGAGATTACATTGATCCAGCTTTGCTCAAAA
AATTCACCAAAGAAACGGGCATTGAAGTGCAGTATGAAACTTTCGATTCCAATGAAGCCATGTACACTAAAATCAAGCAG
GGCGGAACCACTTACGACATTGCTGTTCCTAGTGATTACACCATTGATAAAATGATCAAAGAAAACCTACTCAATAAGCT
TGATAAGTCAAAATTAGTTGGCATGGATAATATCGGGAAAGAATTTTTAGGGAAAAGCTTTGACCCACAAAACGACTATT
CTTTGCCTTATTTCTGGGGAACCGTTGGGATTGTTTATAATGATCAATTAGTTGATAAGGCGCCTATGCACTGGGAAGAT
CTGTGGCGTCCAGAATATAAAAATAGTATTATGCTGATTGATGGAGCGCGTGAAATGCTAGGGGTTGGTTTAACAACTTT
TGGTTATAGTGTGAATTCTAAAAATCTAGAGCAGTTGCAGGCAGCCGAGAGAAAACTGCAGCAGTTGACGCCGAATGTTA
AAGCCATTGTAGCAGATGAGATGAAAGGCTACATGATTCAAGGTGACGCTGCTATTGGAATTACCTTTTCTGGTGAAGCC
AGTGAGATGTTAGATAGTAACGAACACCTTCACTACATCGTGCCTTCAGAAGGGTCTAACCTTTGGTTTGATAATTTGGT
ACTACCAAAAACCATGAAACACGAAAAAGAAGCTTATGCTTTTTTGAACTTTATCAATCGTCCTGAAAATGCTGCGCAAA
ATGCTGCATATATTGGTTATGCGACACCAAATAAAAAAGCCAAGGCCTTACTTCCAGATGAGATAAAAAATGATCCTGCT
TTTTATCCAACAGATGACATTATCAAAAAATTGGAAGTTTATGACAATTTAGGGTCAAGATGGTTGGGGATTTATAATGA
TTTATACCTCCAATTTAAAATGTATCGCAAATAA.

Preferred GAS 159 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 61; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 61, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS 159 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 61. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 61. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 61. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 61 (shown below) is removed. (SEQ ID NO: 63 comprises the N-terminus leader amino acid sequence. SEQ ID NO: 64 comprises a fragment of GAS 159 where the N-terminus leader amino acid sequence is removed). In another example, the underlined amino acid sequence at the C-terminus of SEQ ID NO: 61 is removed. (SEQ ID NO: 65 comprises the C-terminus hydrophobic region. SEQ ID NO: 66 comprises a fragment of GAS 159 where the C-terminus hydrophobic region is removed. SEQ ID NO: 67 comprises a fragment of GAS 159 where both the N-terminus leader sequence and C-terminus hydrophobic region are removed.) Other fragments omit one or more domains of

the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 61

MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSNEAMYTKIKQ
GGTTYDIAVPSDYTIDKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPQNDYSLPYFWGTVGIVYNDQLVDKAPMHWED
LWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNVKAIVADEMKGYMIQGDAAIGITFSGEA
SEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFLNFINRPENAAQNAAYIGYATPNKKAKALLPDEIKNDPA
FYPTDDIIKKLEVYDNLGSRWLGIYNDLYLQFKMYRK

(11) GAS 217

GAS 217 corresponds to M1 GenBank accession numbers GI:13622089 and GI:15674945, to M3 GenBank accession number GI: 21910174, to M18 GenBank accession number GI: 19745987, and is also referred to as 'Spy0925' (M1), 'SpyM3_0638' (M3), and 'SpyM18_0982' (M18). GAS 217 has also been identified as a putative oxidoreductase. Amino acid and polynucleotide sequences of GAS 217 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 68 and 69.

Preferred GAS 217 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 68; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 68, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These GAS 217 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 68. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 68. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 68. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(12) GAS 309

GAS 309 corresponds to M1 GenBank accession numbers GI:13621426 and GI:15674341, to M3 GenBank accession number GI: 21909633, to M18 GenBank accession number GI: 19745363, and is also referred to as 'Spy0124' (M1), 'SpyM3_0097' (M3), 'SpyM18_0205' (M18), 'nra' and 'rofA'. GAS 309 has also been identified as a regulatory protein and a negative transcriptional regulator. Amino acid and polynucleotide sequences of GAS 309 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 70 and 71.

Preferred GAS 309 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 70; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 70, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These GAS 309 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 70. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 70. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 70. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(13) GAS 372

GAS 372 corresponds to M1 GenBank accession numbers GI:13622698 and GI:15675501, to M3 GenBank accession number GI: 21910905, to M18 GenBank accession number GI: 19746500 and is also referred to as 'Spy1625' (M1), 'SpyM3_1369' (M3), and 'SpyM18_1634' (M18). GAS 372 has also been identified as a putative protein kinase or a putative eukaryotic-type serine/threonine kinase. Amino acid and polynucleotide sequences of GAS 372 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 72 and 73.

Preferred GAS 372 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 72; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 72, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). These GAS 372 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 72. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 72. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 72. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(14) GAS 039

GAS 039 corresponds to M1 GenBank accession numbers GI:13621542 and GI:15674446, to M3 GenBank accession number GI: 21909730, to M18 GenBank accession number GI: 19745398 and is also referred to as 'Spy0266' (M1), 'SpyM3_0194' (M3), and 'SpyM18_0250' (M18). Amino acid and polynucleotide sequences of GAS 039 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 74 and 75.

Preferred GAS 039 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 74; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 74, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, or more). These GAS 039 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 74. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 74. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 74. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(15) GAS 042

GAS 042 corresponds to M1 GenBank accession numbers GI:13621559 and GI:15674461, to M3 GenBank accession number GI: 21909745, to M18 GenBank accession number GI: 19745415, and is also referred to as 'Spy0287' (M1), 'SpyM3_0209' (M3), and 'SpyM18_0275' (M18). Amino acid and polynucleotide sequences of GAS 042 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 76 and 77.

Preferred GAS 042 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 76; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 76, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, or more). These GAS 042 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 76. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 76. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 76. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(16) GAS 058

GAS 058 corresponds to M1 GenBank accession numbers GI:13621663 and GI:15674556, to M3 GenBank accession number GI: 21909841, to M18 GenBank accession number GI: 19745567 and is also referred to as 'Spy0430' (M1), 'SpyM3_0305' (M3), and 'SpyM18_0477' (M18). Amino acid and polynucleotide sequences of GAS 058 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 78 and 79.

Preferred GAS 058 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 78; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 78, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, or more). These GAS 058 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 78. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 78. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 78. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 78 (shown below) is removed. (SEQ ID NO: 80 comprises the N-terminal leader sequence. SEQ ID NO: 81 comprises a fragment of GAS 58 where the N-terminal leader sequence is removed.) Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 78

MKWSGFMKTKSKRFLNLATLCLALLGTTLLMAHPVQAEVISKRDYMTRFGLGDLEDDSANYPSNLEARYKGYLEGYEKGL
KGDDIPERPKIQVPEDVQPSDHGDYRDGYEEGFGEGQHKRDPLETEAEDDSQGGRQEGRQGHQEGADSSDLNVEESDGLS
VIDEVVGVIYQAFSTIWTYLSGLF

(17) GAS 290

GAS 290 corresponds to M1 GenBank accession numbers GI:13622978 and GI:15675757, to M3 GenBank accession number GI: 21911221, to M18 GenBank accession number GI: 19746869 and is also referred to as 'Spy1959' (M1), 'SpyM3_1685' (M3), and 'SpyM18_2026' (M18). Amino acid and polynucleotide sequences of GAS 290 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 82 and 83.

Preferred GAS 290 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 82; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 82, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). These GAS 290 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 82. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 82. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 82. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(18) GAS 511

GAS 511 corresponds to M1 GenBank accession numbers GI:13622798 and GI:15675592, to M3 GenBank accession number GI: 21911053, to M18 GenBank accession number GI: 19746700 and is also referred to as 'Spy1743' (M1), 'SpyM3_1517' (M3), 'SpyM18_1815' (M18) and 'accA'. Amino acid and polynucleotide sequences of GAS 511 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 84 and 85.

Preferred GAS 511 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 84; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 84, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). These GAS 511 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 84. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 84. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 84. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(19) GAS 533

GAS 533 corresponds to M1 GenBank accession numbers GI:13622912 and GI:15675696, to M3 GenBank accession number GI: 21911157, to M18 GenBank accession number GI: 19746804 and is also referred to as 'Spy1877' (M1), 'SpyM3_1621' (M3), 'SpyM18_1942' (M18) and 'glnA'. GAS 533 has also been identified as a putative glutamine synthetase. Amino acid and polynucleotide sequences of GAS 533 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 86 and 87.

Preferred GAS 533 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 86; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 86, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 533 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 86. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 86. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20,

25 or more) from the N-terminus of SEQ ID NO: 86. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(20) GAS 527

GAS 527 corresponds to M1 GenBank accession numbers GI:13622332, GI:15675169, and GI:24211764, to M3 GenBank accession number GI: 21910381, to M18 GenBank accession number GI: 19746136, and is also referred to as 'Spy1204' (M1), 'SpyM3_0845' (M3), 'SpyM18_1155' (M18) and 'guaA'. GAS 527 has also been identified as a putative GMP synthetase (glutamate hydrolyzing) (glutamate amidotransferase). Amino acid and polynucleotide sequences of GAS 527 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 88 and 89.

Preferred GAS 527 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 88; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 88, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 527 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 88. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 88. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 88. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(21) GAS 294

GAS 294 corresponds to M1 GenBank accession numbers GI:13622306, GI:15675145, and GI:26006773, to M3 GenBank accession number GI: 21910357, to M18 GenBank accession number GI: 19746111 and is also referred to as 'Spy1173' (M1), 'SpyM3_0821' (M3), 'SpyM18_1125' (M18) and 'gid'. GAS 294 has also been identified as a putative glucose-inhibited division protein. Amino acid and polynucleotide sequences of GAS 294 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 90 and 91.

Preferred GAS 294 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 90; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 90, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 294 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 90. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 90. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 90. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(22) GAS 253

GAS 253 corresponds to M1 GenBank accession numbers GI:13622611, GI:15675423, and GI:21362716, to M3 GenBank accession number GI: 21910711, to M18 GenBank accession number GI: 19746473 and is also referred to as 'Spy1524' (M1), 'SpyM3_1175' (M3), 'SpyM18_1541' (M18) and 'murG'. GAS 253 has also been identified as a putative undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc transferase. Amino acid and polynucleotide sequences of GAS 253 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 92 and 93.

Preferred GAS 253 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 92; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 92, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 253 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 92. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 92. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 92. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(23) GAS 529

GAS 529 corresponds to M1 GenBank accession numbers GI:13622403, GI:15675233, and GI:21759132, to M3 GenBank accession number GI: 21910446, to M18 GenBank accession number GI: 19746203 and is also referred to as 'Spy1280' (M1), 'SpyM3_0910' (M3), 'SpyM18_1228' (M18) and 'glmS'. GAS 529 has also been identified as a putative L-glutamine-D-fructose-6-phosphate aminotransferase (Glucosamine-6-phophate synthase). Amino acid and polynucleotide sequences of GAS 529 of an M1 strain are set forth below and in the sequence listing as SEQ ID NOS: 94 and 95.

Preferred GAS 529 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 94; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 94, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 529 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 94. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 94. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 94. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(24) GAS 045

GAS 045 corresponds to M3 GenBank accession number GI: 21909751, M18 GenBank accesion number GI: 19745421 and is referred to as 'SpyM3_0215' (M3), 'SpyM18_oppA' (M18) and 'oppA'. GAS 045 has been identified as an oligopeptide permease. Amino acid and polynucleotide sequences of GAS 045 from an M1 strain are set forth in the sequence listing as SEQ ID NOS: 96 and 97.

Preferred GAS 045 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 96; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 96, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 045 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 96. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 96. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 96. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 96 (shown below) is removed. (SEQ ID NO: 98 comprises the underlined N-terminal leader sequence. SEQ ID NO: 99 comprises a fragment of GAS 45 where the N-terminal leader sequence is removed). Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 96

VTFMKKSKWLAAVSVAILSVSALAACGNKNASGGSEATKTYKYVFVNDPKSLDYILTNGGGTTDVITQMVDGLLENDEYG
NLVPSLAKDWKVSKDGLTYTYTLRDGVSWYTADGEEYAPVTAEDFVTGLKHAVDDKSDALYVVEDSIKNLKAYQNGEVDF
KEVGVKALDDKTVQYTLNKPESYWNSKTTYSVLFPVNAKFLKSKGKDFGTTDPSSILVNGAYFLSAFTSKSSMEFHKNEN
YWDAKNVGIESVKLTYSDGSDPGSFYKNFDKGEFSVARLYPNDPTYKSAKKNYADNITYGMLTGDIRHLTWNLNRTSFKN
TKKDPAQQDAGKKALNNKDFRQAIQFAFDRASFQAQTAGQDAKTKALRNMLVPPTFVTIGESDFGSEVEKEMAKLGDEWK
DVNLADAQDGFYNPEKAKAEFAKAKEALTAEGVTFPVQLDYPVDQANAATVQEAQSFKQSVEASLGKENVIVNVLETETS
THEAQGFYAETPEQQDYDIISSWWGPDYQDPRTYLDIMSPVGGGSVIQKLGIKAGQNKDVVAAAGLDTYQTLLDEAAAIT
DDNDARYKAYAKAQAYLTDNAVDIPVVALGGTPRVTKAVPFSGGFSWAGSKGPLAYKGMKLQDKPVTVKQYEKAKEKWMK
AKAKSNAKYAEKLADHVEK

(25) GAS 095

GAS 095 corresponds to M1 GenBank accession numbers GI:13622787 and GI:15675582, to M3 GenBank accession number GI: 21911042, to M18 GenBank accession number GI: 19746634 and is also referred to as 'Spy1733' (M1), 'SpyM3_1506' (M3), 'SpyM18_1741' (M18). GAS 095 has also been identified as a putative transcription regulator. Amino acid and polynucleotide sequences of GAS 095 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 100 and 101.

Preferred GAS 095 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 100; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 100, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 095 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 100. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 100. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15,

20, 25 or more) from the N-terminus of SEQ ID NO: 100. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 100 (shown below) is removed. (SEQ ID NO: 102 comprises the amino acid sequence of the underlined N-terminal leader sequence. SEQ ID NO: 103 comprises a fragment of GAS 95 where the N-terminal leader sequence is removed.) Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 100

MKIGKKIVLMFTAIVLTTVLALGVYLTSAYTFSTGELSKTFKDFSTSSNKSDAIKQTRAFSILLMGVDTGSSERASKWEG
NSDSMILVTVNPKTKKTTMTSLERDTLTTLSGPKNNEMNGVEAKLNAAYAAGGAQMAIMTVQDLLNITIDNYVQINMQGL
IDLVNAVGGITVTNEFDFPISIAENEPEYQATVAPGTHKINGEQALVYARMRYDDPEGDYGRQKRQREVIQKVLKKILAL
DSISSYRKILSAVSSNMQTNIEISSRTIPSLLGYRDALRTIKTYQLKGEDATLSDGGSYQIVTSNHLLEIQNRIRTELGL
HKVNQLKTNATVYENLYGSTKSQTVNNNYDSSGQAPSYSDSHSSYANYSSGVDTGQSASTDQDSTASSHRPATPSSSSDA
LAADESSSSGSGSLVPPANINPQT

(26) GAS 193

GAS 193 corresponds to M1 GenBank accession numbers GI:13623029 and GI:15675802, to M3 GenBank accession number GI: 21911267, to M18 GenBank accession number GI: 19746914 and is also referred to as 'Spy2025' (M1), 'SpyM3_1731' (M3), 'SpyM18_2082' (M18) and 'isp'. GAS 193 has also been identified as an immunogenic secreted protein precursor. Amino acid and polynucleotide sequences of GAS 193 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 104 and 105.

Preferred GAS 193 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 104; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 104, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 193 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 104. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 104. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 104. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(27) GAS 137

GAS 137 corresponds to M1 GenBank accession numbers GI:13621842, GI:15674720 and GI:30173478, to M3 GenBank accession number GI:21909998, to M18 GenBank accession number GI: 19745749 and is also referred to as 'Spy0652' (M1), 'SpyM3_0462', and 'SpyM18_0713' (M18). Amino acid and polynucleotide sequences of GAS 137 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 106 and 107.

Preferred GAS 137 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 106; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 106, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 137 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 106. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 106. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,

15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 106. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(28) GAS 084

GAS 084 corresponds to M1 GenBank accession numbers GI:13622398 and GI:15675229, to M3 GenBank accession number GI: 21910442, to M18 GenBank accession number GI: 19746199 and is also referred to as 'Spy1274' (M1), 'SpyM3_0906' and 'SpyM18_1223' (M18). GAS 084 has also been identified as a putative amino acid ABC transporter/periplasmic amino acid binding protein. Amino acid and polynucleotide sequences of GAS 084 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 108 and 109.

Preferred GAS 084 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 108; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 108, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 084 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 108. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 108. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 108. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 108 (shown below) is removed. (SEQ ID NO: 110 comprises an amino acid sequence comprising the underlined N-terminal leader sequence of GAS 84. SEQ ID NO: 111 comprises a fragment of GAS 84 where the N-terminal leader sequence is removed). Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 108

MIIKKRTVAILAIASSFFLVACQATKSLKSGDAWGVYQKQKSITVGFDNTFVPMGYKDESGRCKGFDIDLAKEVFHQYGL
KVNFQAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAFTDSYMRNEQIIVVKKRSDIKTISDMKHKVLGAQSASSGYD
SLLRTPKLLKDFIKNKDANQYETFTQAFIDLKSDRIDGILIDKVYANYYLAKEGQLENYRMIPTTFENEAFSVGLRKEDK
TLQAKINRAFRVLYQNGKFQAISEKWFGDDVATANIKS

(29) GAS 384

GAS 384 corresponds to M1 GenBank accession numbers GI:13622908 and GI:15675693, to M3 GenBank accession number GI: 21911154, to M18 GenBank accession number GI: 19746801 and is also referred to as 'Spy1874' (M1), 'SpyM3_1618' (M3), and 'SpyM18_1939' (M18). GAS 384 has also been identified as a putative glycoprotein endopeptidase. Amino acid and polynucleotide sequences of GAS 384 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 112 and 113.

Preferred GAS 384 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 112; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 112, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 384 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 112. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 112. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,

15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 112. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(30) GAS 202

GAS 202 corresponds to M1 GenBank accession numbers GI:13622431 and GI:15675258, to M3 GenBank accession number GI: 21910527, to M18 GenBank accession number GI: 19746290 and is also referred to as 'Spy1309' (M1), 'SpyM3_0991' (M3), 'SpyM18_1321' (M18) and 'dltD'. GAS 202 has also been identified as a putative extramembranal protein. Amino acid and polynucleotide sequences of GAS 202 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 114 and 115.

Preferred GAS 202 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 114; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 114, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 202 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 114. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 114. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 114. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(31) GAS 057

GAS 057 corresponds to M1 GenBank accession numbers GI:13621655 and GI:15674549, to M3 GenBank accession number GI: 21909834, to M18 GenBank accession number GI: 19745560 and is also referred to as 'Spy0416' (M1), 'SpyM3_0298' (M3), 'SpyM18_0464' (M18) and 'prtS'. GAS 057 has also been identified as a putative cell envelope proteinase. Amino acid and polynucleotide sequences of GAS 057 of an M1 strain are set forth in the sequence listing as SEQ ID NOS: 116 and 117.

Preferred GAS 057 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 116; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 116, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 057 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 116. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 116. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 116. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 116 (shown below) is removed. (SEQ ID NO: 118 comprises the underlined N-terminal leader sequence. SEQ ID NO: 119 comprises a fragment of GAS 57 where the N-terminal leader sequence is removed.) In another example, the underlined amino acid sequence at the C-terminus of SEQ ID NO: 116 is removed. (SEQ ID NO: 120 comprises the underlined C-terminal hydrophobic region. SEQ ID NO: 121 comprises a fragment of GAS 57 where the C-terminal hydrophobic

region is removed. SEQ ID NO: 122 comprises a fragment of GAS 57 where both the N-terminal leader sequence and the C-terminal hydrophobic region are removed.) Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

SEQ ID NO: 116

MEKKQRFSLRKYKSGTFSVLIGSVFLVMTTTVAADELSTMSEPTITNHAQQQAQHLTNTELSSAESKSQDTSQITLKTNR
EKEQSQDLVSEPTTTELADTDAASMANTGSDATQKSASLPPVNTDVHDWVKTKGAWDKGYKGQGKVVAVIDTGIDPAHQS
MRISDVSTAKVKSKEDMLARQKAAGINYGSWINDKVVFAHNYVENSDNIKENQFEDFDEDWENFEFDAEAEPKAIKKHKI
YRPQSTQAPKETVIKTEETDGSHDIDWTQTDDDTKYESHGMHVTGIVAGNSKEAAATGERFLGIAPEAQVMFMRVFANDI
MGSAESLFIKAIEDAVALGADVINLSLGTANGAQLSGSKPLMEAIEKAKKAGVSVVVAAGNERVYGSDHDDPLATNPDYG
LVGSPSTGRTPTSVAAINSKWVIQRLMTVKELENRADLNHGKAIYSESVDFKDIKDSLGYDKSHQFAYVKESTDAGYNAQ
DVKGKIALIERDPNKTYDEMIALAKKHGALGVLIFNNKPGQSNRSMRLTANGMGIPSAFISHEFGKAMSQLNGNGTGSLE
FDSVVSKAPSQKGNEMNHFSNWGLTSDGYLKPDITAPGGDIYSTYNDNHYGSQTGTSMASPQIAGASLLVKQYLEKTQPN
LPKEKIADIVKNLLMSNAQIHVNPETKTTTSPRQQGAGLLNIDGAVTSGLYVTGKDNYGSISLGNITDTMTFDVTVHNLS
NKDKTLRYDTELLTDHVDPQKGRFTLTSHSLKTYQGGEVTVPANGKVTVRVTMDVSQFTKELTKQMPNGYYLEGFVRFRD
SQDDQLNRVNIPFVGFKGQFENLAVAEESIYRLKSQGKTGFYFDESGPKDDIYVGKHFTGLVTLGSETNVSTKTISDNGL
HTLGTFKNADGKFILEKNAQGNPVLAISPNGDNNQDFAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVSPESFKGDKNF
NSDIRFAKSTTLLGTAFSGKSLTGAELPDGHYHYVVSYYPDVVGAKRQEMTFDMILDRQKPVLSQATFDPETNRFKPEPL
KDRGLAGVRKDSVFYLERKDNKPYTVTINDSYKYVSVEDNKTFVERQADGSFILPLDKAKLGDFYYMVEDFAGNVAIAKL
GDHLPQTLGKTPIKLKLTDGNYQTKETLKDNLEMTQSDTGLVTNQAQLAVVHRNQPQSQLTKMNQDFFISPNEDGNKDFV
AFKGLKNNVYNDLTVNVYAKDDHQKQTPIWSSQAGASVSAIESTAWYGITARGSKVMPGDYQYVVTYRDEHGKEHQKQYT
ISVNDKKPMITQGRFDTINGVDHFTPDKTKALDSSGIVREEVFYLAKKNGRKFDVTEGKDGITVSDNKVYIPKNPDGSYT
ISKRDGVTLSDYYYLVEDRAGNVSFATLRDLKAVGKDKAVVNFGLDLPVPEDKQIVNFTYLVRDADGKPIENLEYYNNSG
NSLILPYGKYTVELLTYDTNAAKLESDKIVSFTLSADNNFQQVTFKITMLATSQITAHFDHLLPEGSRVSLKTAQDQLIP
LEQSLYVPKAYGKTVQEGTYEVVVSLPKGYRIEGNTKVNTLPNEVHELSLRLVKVGDASDSTGDHKVMSKNNSQALTASA
TPTKSTTSATAKALPSTGEKMGLKLRIVGLVLLGLTCVFSRKKSTKD

Representative examples of immunization with GAS antigens of the invention in the murine mouse model discussed above are summarized in Figure 8. The first column identifies the GAS antigen used in the experiment. In some instances purification aspects are referenced in this list. Also, modifications to the polynucleotide sequence which have been made to facilitate the recombinant expression of the antigen are denoted in the chart with the following annotations: "a" indicates that N or C terminal hydrophobic regions have been removed; RR indicates codon optimisation; "NH" and "CH" correspond to the expression vectors similar to those indicated in the GAS 40 construct examples. Where a p value is given, it was calculated based on the control HIS stop values at the bottom of the chart.

Mice immunized with GAS 40 yielded substantially improved survival rates on challenge – in a collection of over 100 mice immunizations, immunization with GAS 40 yielded over 50% survival. The other GAS antigens in the chart offered an amount of protection that, for example if combined with GAS 40, could offer improved protection.

The immunogenicity of other known GAS antigens may be improved by combination with two or more GAS the first antigen group. Such other known GAS antigens include a second antigen group consisting of (1) one or more variants of the M surface protein or fragments thereof, (2) fibronectin-binding protein, (3) streptococcal heme-associated protein, or (4) SagA. These antigens are referred to herein as the "second antigen group".

The invention thus includes an immunogenic composition comprising a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of the first antigen group and one, two, three, or four GAS antigens of the second antigen group. Preferably, the combination consists of three, four, five, six, seven, eight, nine, or ten GAS antigens from the first antigen group. Still more preferably, the

combination consists of three, four or five GAS antigens from the first antigen group. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117. Preferably, the combination of GAS antigens includes one or more variants of the M surface protein.

Each of the GAS antigens of the second antigen group are described in more detail below.

(1) M surface protein

The M protein is a GAS virulence factor which has been associated with both colonization and resistance to phagocytosis. Over 100 different type variants of the M protein have been identified on the basis of antigenic specificity and M protein is thought to be the major cause of antigenic shift and antigenic drift in GAS. The M protein also binds fibrinogen from serum and blocks the binding of complement to the underlying peptidoglycan. This action is thought to increase GAS survival within a mammalian host by inhibiting phagocytosis.

Unfortunately, the GAS M protein contains some epitopes which mimic those of mammalian muscle and connective tissue. Certain GAS M proteins may be rheumatogenic since they contain epitopes related to heart muscle, and may lead to autoimmune rheumatic carditis (rheumatic fever) following an acute infection.

Epitopes having increased bactericidal activity and having decreased likelihood of cross-reacting with human tissues have been identified in the amino terminal region and combined into fusion proteins containing approximately six, seven, or eight M protein fragments linked in tandem. See Hu et al., Infection & Immunity (2002) 70(4):2171 – 2177; Dale, Vaccine (1999) 17:193 – 200; Dale et al., Vaccine 14(10):944 – 948; WO 02/094851 and WO 94/06465. (Each of the M protein variants, fragments and fusion proteins described in these references are specifically incorporated herein by reference.)

Accordingly, the compositions of the invention may further comprise a GAS M surface protein or a fragment or derivative thereof. One or more GAS M surface protein fragments may be combined together in a fusion protein. Alternatively, one or more GAS M surface protein fragments are combined with a GAS antigen or fragment thereof of the first antigen group. One example of a GAS M protein is set forth in the sequence listing as SEQ ID NO: 123.

Preferred GAS M proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to a known M protein such as SEQ ID NO: 123; and/or (b) which is a fragment of at least *n* consecutive amino acids of a known M protein such as SEQ ID NO: 123, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS M proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 123. Preferred fragments of (b) comprise an epitope from a known M protein, such as SEQ ID NO: 123. Preferably, the fragment is one of those described in the references above. Preferably, the fragment is constructed in a fusion protein with one or more additional M protein fragments. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of a known M protein such as SEQ ID NO: 123. Other fragments omit one or more domains of the protein (*e.g.*

omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(2) Fibronectin-binding protein

GAS fibronectin-binding protein ('SfbI') is a mutlifunctional bacterial protein thought to mediate attachment of the bacteria to host cells, facilitate bacterial internalization into cells and to bind to the Fc fragment of human IgG, thus interfering with Fc-receptor mediated phagocytosis and antibody-dependent cell cytotoxicity. Immunization of mice with SfbI and an 'H12 fragment' (encoded by positions 1240 – 1854 of the SfbI gene) are discussed in Schulze et al., Vaccine (2003) 21:1958 – 1964; Schulze et al., Infection and Immunity (2001) 69(1):622 – 625 and Guzman et al., Journal of Infectious Diseases (1999) 179:901 – 906. One example of an amino acid sequence for GAS SfbI is shown in the sequence listing as SEQ ID NO: 124.

Preferred SfbI proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 124; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 124, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These SfbI proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 124. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 124. Preferably, the fragment is one of those described in the references above. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 124. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(3) Streptococcal heme-associated protein

The GAS streptococcal heme-associated protein ('Shp') has been identified as a GAS cell surface protein. It is thought to be cotrascribed with genes encoding homologues of an ABC transporter involved in iron uptake in gram-negative bacteria. The Shp protein is further described in Lei et al., "Identification and Characterization of a Novel Heme-Associated Cell Surface Protein Made by *Streptococcus pyogenes*", Infection and Immunity (2002) 70(8):4494 – 4500. One example of a Shp protein is shown in the sequence listing as SEQ ID NO: 125.

Preferred Shp proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 125; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 125, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). These Shp proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 125. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 125. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 125. Other fragments omit one or more domains of the protein

(*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

(4) SagA

Streptolysin S (SLS), also known as 'SagA', is thought to be produced by almost all GAS colonies. This cytolytic toxin is responsible for the beta-hemolysis surrounding colonies of GAS grown on blood agar and is thought to be associated with virulence. While the full SagA peptide has not been shown to be immunogenic, a fragment of amino acids 10 – 30 (SagA 10 – 30) has been used to produce neutralizing antibodies. See Dale et al., "Antibodies against a Synthetic Peptide of SagA Neutralize the Cytolytic Activity of Streptolysin S from Group A Streptococci", Infection and Immunity (2002) 70(4):2166 – 2170. The amino acid sequence of SagA 10 – 30 is shown in the sequence listing as SEQ ID NO: 126.

Preferred SagA 10-30 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 126; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 126, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, or 20). These SagA 10 - 30 proteins include variants (*e.g.* allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 126.

There is an upper limit to the number of GAS antigens which will be in the compositions of the invention. Preferably, the number of GAS antigens in a composition of the invention is less than 20, less than 19, less than 18, less than 17, less than 16, less than 15, less than 14, less than 13, less than 12, less than 11, less than 10, less than 9, less than 8, less than 7, less than 6, less than 5, less than 4, or less than 3. Still more preferably, the number of GAS antigens in a composition of the invention is less than 6, less than 5, or less than 4. Still more preferably, the number of GAS antigens in a composition of the invention is 3. The GAS antigens used in the invention are preferably isolated, i.e., separate and discrete, from the whole organism with which the molecule is found in nature or, when the polynucleotide or polypeptide is not found in nature, is sufficiently free of other biological macromolecules so that the polynucleotide or polypeptide can be used for its intended purpose.

Fusion proteins

The GAS antigens used in the invention may be present in the composition as individual separate polypeptides, but it is preferred that at least two (*i.e.* 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20) of the antigens are expressed as a single polypeptide chain (a 'hybrid' polypeptide). Hybrid polypeptides offer two principal advantages: first, a polypeptide that may be unstable or poorly expressed on its own can be assisted by adding a suitable hybrid partner that overcomes the problem; second, commercial manufacture is simplified as only one expression and purification need be employed in order to produce two polypeptides which are both antigenically useful.

The hybrid polypeptide may comprise two or more polypeptide sequences from the first antigen group. Accordingly, the invention includes a composition comprising a first amino acid sequence and a second amino acid sequence, wherein said first and second amino acid sequences are selected from a GAS

antigen or a fragment thereof of the first antigen group. Preferably, the first and second amino acid sequences in the hybrid polypeptide comprise different epitopes.

The hybrid polypeptide may comprise one or more polypeptide sequences from the first antigen group and one or more polypeptide sequences from the second antigen group. Accordingly, the invention includes a composition comprising a first amino acid sequence and a second amino acid sequence, said first amino acid sequence selected from a GAS antigen or a fragment thereof from the first antigen group and said second amino acid sequence selected from a GAS antigen or a fragment thereof from the second antigen group. Preferably, the first and second amino acid sequences in the hybrid polypeptide comprise different epitopes.

Hybrids consisting of amino acid sequences from two, three, four, five, six, seven, eight, nine, or ten GAS antigens are preferred. In particular, hybrids consisting of amino acid sequences from two, three, four, or five GAS antigens are preferred.

Different hybrid polypeptides may be mixed together in a single formulation. Within such combinations, a GAS antigen may be present in more than one hybrid polypeptide and/or as a non-hybrid polypeptide. It is preferred, however, that an antigen is present either as a hybrid or as a non-hybrid, but not as both.

Hybrid polypeptides can be represented by the formula **NH_2-A-{-X-L-}$_n$-B-COOH**, wherein: X is an amino acid sequence of a GAS antigen or a fragment thereof from the first antigen group or the second antigen group; L is an optional linker amino acid sequence; A is an optional N-terminal amino acid sequence; B is an optional C-terminal amino acid sequence; and *n* is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

If a -X- moiety has a leader peptide sequence in its wild-type form, this may be included or omitted in the hybrid protein. In some embodiments, the leader peptides will be deleted except for that of the -X- moiety located at the N-terminus of the hybrid protein *i.e.* the leader peptide of X_1 will be retained, but the leader peptides of X_2 ... X_n will be omitted. This is equivalent to deleting all leader peptides and using the leader peptide of X_1 as moiety -A-.

For each *n* instances of {-X-L-}, linker amino acid sequence -L- may be present or absent. For instance, when *n*=2 the hybrid may be NH_2-X_1-L_1-X_2-L_2-COOH, NH_2-X_1-X_2-COOH, NH_2-X_1-L_1-X_2-COOH, NH_2-X_1-X_2-L_2-COOH, *etc.* Linker amino acid sequence(s) -L- will typically be short (*e.g.* 20 or fewer amino acids *i.e.* 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples comprise short peptide sequences which facilitate cloning, poly-glycine linkers (*i.e.* comprising Gly_n where $n = 2, 3, 4, 5, 6, 7, 8, 9, 10$ or more), and histidine tags (*i.e.* His_n where $n = 3, 4, 5, 6, 7, 8, 9, 10$ or more). Other suitable linker amino acid sequences will be apparent to those skilled in the art. A useful linker is GSGGGG, with the Gly-Ser dipeptide being formed from a *Bam*HI restriction site, thus aiding cloning and manipulation, and the $(Gly)_4$ tetrapeptide being a typical poly-glycine linker.

-A- is an optional N-terminal amino acid sequence. This will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include leader sequences to direct protein

trafficking, or short peptide sequences which facilitate cloning or purification (*e.g.* histidine tags *i.e.* His_n where $n = 3, 4, 5, 6, 7, 8, 9, 10$ or more). Other suitable N-terminal amino acid sequences will be apparent to those skilled in the art. If X_1 lacks its own N-terminus methionine, -A- is preferably an oligopeptide (*e.g.* with 1, 2, 3, 4, 5, 6, 7 or 8 amino acids) which provides a N-terminus methionine.

-B- is an optional C-terminal amino acid sequence. This will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include sequences to direct protein trafficking, short peptide sequences which facilitate cloning or purification (*e.g.* comprising histidine tags *i.e.* His_n where $n =$ 3, 4, 5, 6, 7, 8, 9, 10 or more), or sequences which enhance protein stability. Other suitable C-terminal amino acid sequences will be apparent to those skilled in the art.

Most preferably, n is 2 or 3.

The fusion constructs of the invention may include a combination of two or more GAS antigens, wherein said combination includes GAS 40 or a fragment thereof or a polypeptide having sequence identity thereto.

The fusion constructs of the invention may include a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of the first antigen group, said first antigen group consisting of: GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057. Preferably, the combination of GAS antigens consists of three, four, five, six, seven, eight, nine, or ten GAS antigens selected from the first antigen group. Preferably, the combination of GAS antigens consists of three, four, or five GAS antigens selected from the first antigen group.

GAS 39, GAS 40, GAS 57, GAS 117, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511 are particularly preferred GAS antigens for use in the fusion constructs of the invention. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117. Preferably, the combination includes GAS 40.

Recombinant expression of the fusion constructs of the invention may be improved or optimised by the same methods described for the expression of the GAS antigens alone (discussed above). Fusion constructs of GAS 40 and GAS 117 are exemplified below.

In the first example, GAS 117 is linked to GAS 40a-RR. (As discussed above, GAS 40a-RR is a codon optimised GAS 40 sequence where the N-terminal leader sequence and the C-terminal transmembrane sequence are removed). In this construct a GAS 117 fragment (where the N-terminal leader sequence is removed) is placed to the N-terminus of the GAS 40 sequence and a HIS tag is added to the C-terminus of the GAS 40 sequence. This construct is designated "117-40a-RR". Amino acid and polynucleotide sequences for this construct are shown in the sequence listing as SEQ ID NOS: 127 and 128.

The GAS 117 and GAS 40 sequences are preferably linked by a linker sequence comprising multiple Glycine residues. For example, the linker used in 117-40a-RR fusion construct, a linker sequence of SEQ ID NO: 129 (YASGGGS) is used.

In a second example, the relative locations of the GAS 40 and GAS 117 sequences can be exchanged. In this construct, designated "40a-RR-117", the GAS 40a-RR sequence is placed to the N-terminus of the GAS 117 sequence and the HIS tag is added to the C-terminus of the GAS 117 sequence. Amino acid and polynucleotide sequences for this fusion construct are shown in the sequence listing as SEQ ID NOS: 130 and 131.

Alternatively, the fusion constructs may be designed without codon optimisations. For example, polynucleotide and amino acid sequences for fusion construct "117-40a" is shown in the sequence listing as SEQ ID NOS: 132 and 133. (While no codon optimisations were used, three point mutations apparently occurred during the cloning, only one of which involved a conservative amino acid change (Glucine to Glycine). In the murine immunization model (previously discussed above), immunization with "117-40a" has yielded up to 80 % survival upon challenge.

A preferred GAS40 fusion sequence comprises a fragment of GAS 40 comprising one or more of the coiled-coil regions. For example, the fusion construct may comprise a GAS 40 sequence comprising the first coiled-coil region. "117-40N" is an example of this type of construct. Amino acid and polynucleotide sequences for this construct are shown in the sequence listing as SEQ ID NOS; 132 and 133.

The invention also provides nucleic acids encoding hybrid polypeptides of the invention. Furthermore, the invention provides nucleic acid which can hybridise to this nucleic acid, preferably under "high stringency" conditions (*e.g.* 65°C in a 0.1xSSC, 0.5% SDS solution).

The GAS antigens of the invention may also be used to prepare antibodies specific to the GAS antigens. The antibodies are preferably specific to the first or second coiled-coil regions of GAS 40. The invention also includes the use of combination of two or more types of antibodies selected from the group consisting of antibodies specific to GBS 80, GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057. Preferably, the combination includes an antibody specific to GAS 40, or a fragment thereof.

The GAS specific antibodies of the invention include one or more biological moieties that, through chemical or physical means, can bind to or associate with an epitope of a GAS polypeptide. The antibodies of the invention include antibodies which specifically bind to a GAS antigen, preferably GAS 80. The invention includes antibodies obtained from both polyclonal and monoclonal preparations, as well as the following: hybrid (chimeric) antibody molecules (see, for example, Winter *et al.* (1991) *Nature* 349: 293-299; and US Patent No. 4,816,567; $F(ab')_2$ and F(ab) fragments; F_v molecules (non-covalent heterodimers, see, for example, Inbar *et al.* (1972) *Proc Natl Acad Sci USA* 69:2659-2662; and Ehrlich *et al.* (1980) *Biochem* 19:4091-4096); single-chain Fv molecules (sFv) (see, for example, Huston *et al.* (1988) *Proc Natl Acad Sci USA* 85:5897-5883); dimeric and trimeric antibody fragment constructs; minibodies (see, *e.g.*, Pack *et al.* (1992) *Biochem* 31:1579-1584; Cumber *et al.* (1992) *J Immunology* 149B: 120-126); humanized antibody molecules (see, for example, Riechmann *et al.* (1988) *Nature* 332:323-327; Verhoeyan *et al.* (1988) *Science* 239:1534-1536; and U.K. Patent Publication No. GB 2,276,169, published 21 September

1994); and, any functional fragments obtained from such molecules, wherein such fragments retain immunological binding properties of the parent antibody molecule. The invention further includes antibodies obtained through non-conventional processes, such as phage display.

Preferably, the GAS specific antibodies of the invention are monoclonal antibodies. Monoclonal antibodies of the invention include an antibody composition having a homogeneous antibody population. Monoclonal antibodies of the invention may be obtained from murine hybridomas, as well as human monoclonal antibodies obtained using human rather than murine hybridomas. See, *e.g.*, Cote, *et al. Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, p 77.

Polypeptides of the invention can be prepared by various means (*e.g.* recombinant expression, purification from cell culture, chemical synthesis, *etc.*) and in various forms (*e.g.* native, fusions, non-glycosylated, lipidated, *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other GAS or host cell proteins).

Nucleic acid according to the invention can be prepared in many ways (*e.g.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself, *etc.*) and can take various forms (*e.g.* single stranded, double stranded, vectors, probes, *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other GAS or host cell nucleic acids).

The term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones (*e.g.* phosphorothioates, *etc.*), and also peptide nucleic acids (PNA), *etc.* The invention includes nucleic acid comprising sequences complementary to those described above (*e.g.* for antisense or probing purposes).

The invention also provides a process for producing a polypeptide of the invention, comprising the step of culturing a host cell transformed with nucleic acid of the invention under conditions which induce polypeptide expression.

The invention provides a process for producing a polypeptide of the invention, comprising the step of synthesising at least part of the polypeptide by chemical means.

The invention provides a process for producing nucleic acid of the invention, comprising the step of amplifying nucleic acid using a primer-based amplification method (*e.g.* PCR).

The invention provides a process for producing nucleic acid of the invention, comprising the step of synthesising at least part of the nucleic acid by chemical means.

Strains

Preferred polypeptides of the invention comprise an amino acid sequence found in an M1, M3 or M18 strain of GAS. The genomic sequence of an M1 GAS strain is reported at Ferretti et al, PNAS (2001) 98(8):4658 – 4663. The genomic sequence of an M3 GAS strain is reported at Beres et al., PNAS (2002) 99(15):10078 – 10083. The genomic sequence of an M18 GAS strain is reported at Smooet et al., PNAS (2002) 99(7):4668 – 4673.

Where hybrid polypeptides are used, the individual antigens within the hybrid (*i.e.* individual -X- moieties) may be from one or more strains. Where n=2, for instance, X_2 may be from the same strain as X_1

or from a different strain. Where n=3, the strains might be (i) $X_1=X_2=X_3$ (ii) $X_1=X_2\neq X_3$ (iii) $X_1\neq X_2=X_3$ (iv) $X_1\neq X_2\neq X_3$ or (v) $X_1=X_3\neq X_2$, *etc.*

Purification and Recombinant Expression

The GAS antigens of the invention may be isolated from a *Streptococcus pyogenes*, or they may be recombinantly produced, for instance, in a heterologous host. Preferably, the GAS antigens are prepared using a heterologous host. The heterologous host may be prokaryotic (*e.g.* a bacterium) or eukaryotic. It is preferably *E.coli*, but other suitable hosts include *Bacillus subtilis, Vibrio cholerae, Salmonella typhi, Salmonella typhimurium, Neisseria lactamica, Neisseria cinerea, Mycobacteria* (*e.g. M.tuberculosis*), yeasts, *etc.*

Recombinant production of polypeptides is facilitated by adding a tag protein to the GAS antigen to be expressed as a fusion protein comprising the tag protein and the GAS antigen. Such tag proteins can facilitate purification, detection and stability of the expressed protein. Tag proteins suitable for use in the invention include a polyarginine tag (Arg-tag), polyhistidine tag (His-tag), FLAG-tag, Strep-tag, c-myc-tag, S-tag, calmodulin-binding peptide, cellulose-binding domain, SBP-tag,, chitin-binding domain, glutathione S-transferase-tag (GST), maltose-binding protein, transcription termination anti-terminiantion factor (NusA), *E. coli* thioredoxin (TrxA) and protein disulfide isomerase I (DsbA). Preferred tag proteins include His-tag and GST. A full discussion on the use of tag proteins can be found at Terpe et al., Appl Microbiol Biotechnol (2003) 60:523 – 533.

After purification, the tag proteins may optionally be removed from the expressed fusion protein, i.e., by specifically tailored enzymatic treatments known in the art. Commonly used proteases include enterokinase, tobacco etch virus (TEV), thrombin, and factor X_a.

Immunogenic compositions and medicaments

Compositions of the invention are preferably immunogenic compositions, and are more preferably vaccine compositions. The pH of the composition is preferably between 6 and 8, preferably about 7. The pH may be maintained by the use of a buffer. The composition may be sterile and/or pyrogen-free. The composition may be isotonic with respect to humans.

Vaccines according to the invention may either be prophylactic (*i.e.* to prevent infection) or therapeutic (*i.e.* to treat infection), but will typically be prophylactic. Accordingly, the invention includes a method for the therapeutic or prophylactic treatment of a *Streptococcus pyogenes* infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of the immunogenic compositions of the invention. Preferably, the immunogenic composition comprises a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of the first antigen group. Preferably, the combination of GAS antigens consists of three, four, five, six, seven, eight, nine, or ten GAS antigens selected from the first antigen group. Preferably, the combination of GAS antigens consists of three, four, or five GAS antigens selected from the first antigen group. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117.

Alternatively, the invention includes an immunogenic composition comprising a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of the first antigen group and

one, two, three, or four GAS antigens of the second antigen group. Preferably, the combination consists of three, four, five, six, seven, eight, nine, or ten GAS antigens from the first antigen group. Still more preferably, the combination consists of three, four or five GAS antigens from the first antigen group. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117. Preferably, the combination of GAS antigens includes one or more variants of the M surface protein.

The invention also provides a composition of the invention for use as a medicament. The medicament is preferably able to raise an immune response in a mammal (*i.e.* it is an immunogenic composition) and is more preferably a vaccine.

The invention also provides the use of the compositions of the invention in the manufacture of a medicament for raising an immune response in a mammal. The medicament is preferably a vaccine. The invention also provides for a kit comprising a first component comprising a combination of GAS antigens. In one embodiment, the combination of GAS antigens consists of a mixture of two to thirty-one GAS antigens selected from the first antigen group. Preferably, the combination consists of three, four, five, six, seven, eight, nine, or ten GAS antigens from the first antigen group. Preferably, the combination consists of three, four, or five GAS antigens from the first antigen group. Preferably, the combination includes either or both of GAS 117 and GAS 040.

In another embodiment, the kit comprises a first component comprising a combination of GAS antigens consisting of a mixture of two to thirty-one GAS antigens of the first antigen group and one, two, three, or four GAS antigens of the second antigen group. Preferably, the combination consists of three, four, five, six, seven, eight, nine, or ten GAS antigens from the first antigen group. Still more preferably, the combination consists of three, four or five GAS antigens from the first antigen group. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117. Preferably, the combination of GAS antigens includes one or more variants of the M surface protein.

The invention also provides a delivery device pre-filled with the immunogenic compositions of the invention.

The invention also provides a method for raising an immune response in a mammal comprising the step of administering an effective amount of a composition of the invention. The immune response is preferably protective and preferably involves antibodies and/or cell-mediated immunity. The method may raise a booster response.

The mammal is preferably a human. Where the vaccine is for prophylactic use, the human is preferably a child (*e.g.* a toddler or infant) or a teenager; where the vaccine is for therapeutic use, the human is preferably a teenager or an adult. A vaccine intended for children may also be administered to adults *e.g.* to assess safety, dosage, immunogenicity, *etc.*

These uses and methods are preferably for the prevention and/or treatment of a disease caused by *Streptococcus pyogenes* (*e.g.* pharyngitis (such as streptococcal sore throat), scarlet fever, impetigo, erysipelas, cellulitis, septicemia, toxic shock syndrome, necrotizing fasciitis (flesh eating disease) and sequelae (such as rheumatic fever and acute glomerulonephritis)). The compositions may also be effective against other streptococcal bacteria.

One way of checking efficacy of therapeutic treatment involves monitoring GAS infection after administration of the composition of the invention. One way of checking efficacy of prophylactic treatment involves monitoring immune responses against the GAS antigens in the compositions of the invention after administration of the composition.

Compositions of the invention will generally be administered directly to a patient. Direct delivery may be accomplished by parenteral injection (*e.g.* subcutaneously, intraperitoneally, intravenously, intramuscularly, or to the interstitial space of a tissue), or by rectal, oral (*e.g.* tablet, spray), vaginal, topical, transdermal (*e.g.* see WO99/27961) or transcutaneous (*e.g.* see WO02/074244 and WO02/064162), intranasal (*e.g.* see WO03/028760), ocular, aural, pulmonary or other mucosal administration. The invention may be used to elicit systemic and/or mucosal immunity.

Dosage treatment can be a single dose schedule or a multiple dose schedule. Multiple doses may be used in a primary immunisation schedule and/or in a booster immunisation schedule. In a multiple dose schedule the various doses may be given by the same or different routes *e.g.* a parenteral prime and mucosal boost, a mucosal prime and parenteral boost, *etc.*

The compositions of the invention may be prepared in various forms. For example, the compositions may be prepared as injectables, either as liquid solutions or suspensions. Solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared (*e.g.* a lyophilised composition). The composition may be prepared for topical administration *e.g.* as an ointment, cream or powder. The composition may be prepared for oral administration *e.g.* as a tablet or capsule, as a spray, or as a syrup (optionally flavoured). The composition may be prepared for pulmonary administration *e.g.* as an inhaler, using a fine powder or a spray. The composition may be prepared as a suppository or pessary. The composition may be prepared for nasal, aural or ocular administration *e.g.* as drops. The composition may be in kit form, designed such that a combined composition is reconstituted just prior to administration to a patient. Such kits may comprise one or more antigens in liquid form and one or more lyophilised antigens.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of antigen(s), as well as any other components, as needed. By 'immunologically effective amount', it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, age, the taxonomic group of individual to be treated (*e.g.* non-human primate, primate, *etc.*), the capacity of the individual's immune system to synthesise antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

Further components of the composition

The composition of the invention will typically, in addition to the components mentioned above, comprise one or more 'pharmaceutically acceptable carriers', which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids,

polyglycolic acids, polymeric amino acids, amino acid copolymers, and lipid aggregates (such as oil droplets or liposomes). Such carriers are well known to those of ordinary skill in the art. The vaccines may also contain diluents, such as water, saline, glycerol, *etc.* Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present. A thorough discussion of pharmaceutically acceptable excipients is available in Gennaro (2000) *Remington: The Science and Practice of Pharmacy.* 20th ed., ISBN: 0683306472.

Vaccines of the invention may be administered in conjunction with other immunoregulatory agents. In particular, compositions will usually include an adjuvant.

Preferred further adjuvants include, but are not limited to, one or more of the following set forth below:

A. Mineral Containing Compositions

Mineral containing compositions suitable for use as adjuvants in the invention include mineral salts, such as aluminium salts and calcium salts. The invention includes mineral salts such as hydroxides (*e.g.* oxyhydroxides), phosphates (*e.g.* hydroxyphoshpates, orthophosphates), sulphates, *etc.* {*e.g.* see chapters 8 & 9 of *Vaccine design:the subunit and adjuvant approach* (1995) Powell & Newman. ISBN 0-306-44867-X}), or mixtures of different mineral compounds, with the compounds taking any suitable form (*e.g.* gel, crystalline, amorphous, *etc.*), and with adsorption being preferred. The mineral containing compositions may also be formulated as a particle of metal salt. See WO00/23105.

B. Oil-Emulsions

Oil-emulsion compositions suitable for use as adjuvants in the invention include squalene-water emulsions, such as MF59 (5% Squalene, 0.5% Tween 80, and 0.5% Span 85, formulated into submicron particles using a microfluidizer). See WO90/14837. See also, Podda, "The adjuvanted influenza vaccines with novel adjuvants: experience with the MF59-adjuvanted vaccine", Vaccine (2001) 19: 2673-2680; Frey et al., "Comparison of the safety, tolerability, and immunogenicity of a MF59-adjuvanted influenza vaccine and a non-adjuvanted influenza vaccine in non-elderly adults", Vaccine (2003) 21:4234-4237. MF59 is used as the adjuvant in the FLUAD™ influenza virus trivalent subunit vaccine.

Particularly preferred adjuvants for use in the compositions are submicron oil-in-water emulsions. Preferred submicron oil-in-water emulsions for use herein are squalene/water emulsions optionally containing varying amounts of MTP-PE, such as a submicron oil-in-water emulsion containing 4-5% w/v squalene, 0.25-1.0% w/v Tween 80 ™ (polyoxyelthylenesorbitan monooleate), and/or 0.25-1.0% Span 85™ (sorbitan trioleate), and, optionally, N-acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-huydroxyphosphophoryloxy)-ethylamine (MTP-PE), for example, the submicron oil-in-water emulsion known as "MF59" (International Publication No. WO90/14837; US Patent Nos. 6,299,884 and 6,451,325, incorporated herein by reference in their entireties; and Ott et al., "MF59 -- Design and Evaluation of a Safe and Potent Adjuvant for Human Vaccines" in *Vaccine Design: The Subunit and Adjuvant Approach* (Powell, M.F. and Newman, M.J. eds.) Plenum Press, New York, 1995, pp. 277-296). MF59 contains 4-5% w/v Squalene (e.g. 4.3%), 0.25-0.5% w/v Tween 80™, and 0.5% w/v Span 85™ and optionally contains various amounts of MTP-PE, formulated into submicron particles using a microfluidizer

such as Model 110Y microfluidizer (Microfluidics, Newton, MA). For example, MTP-PE may be present in an amount of about 0-500 μg/dose, more preferably 0-250 μg/dose and most preferably, 0-100 μg/dose. As used herein, the term "MF59-0" refers to the above submicron oil-in-water emulsion lacking MTP-PE, while the term MF59-MTP denotes a formulation that contains MTP-PE. For instance, "MF59-100" contains 100 μg MTP-PE per dose, and so on. MF69, another submicron oil-in-water emulsion for use herein, contains 4.3% w/v squalene, 0.25% w/v Tween 80™, and 0.75% w/v Span 85™ and optionally MTP-PE. Yet another submicron oil-in-water emulsion is MF75, also known as SAF, containing 10% squalene, 0.4% Tween 80™, 5% pluronic-blocked polymer L121, and thr-MDP, also microfluidized into a submicron emulsion. MF75-MTP denotes an MF75 formulation that includes MTP, such as from 100-400 μg MTP-PE per dose. Submicron oil-in-water emulsions, methods of making the same and immunostimulating agents, such as muramyl peptides, for use in the compositions, are described in detail in International Publication No. WO90/14837 and US Patent Nos. 6,299,884 and 6,45 1,325, incorporated herein by reference in their entireties.

Complete Freund's adjuvant (CFA) and incomplete Freund's adjuvant (IFA) may also be used as adjuvants in the invention.

C. Saponin Formulations

Saponin formulations, may also be used as adjuvants in the invention. Saponins are a heterologous group of sterol glycosides and triterpenoid glycosides that are found in the bark, leaves, stems, roots and even flowers of a wide range of plant species. Saponin from the bark of the *Quillaia saponaria* Molina tree have been widely studied as adjuvants. Saponin can also be commercially obtained from *Smilax ornata* (sarsaprilla), *Gypsophilla paniculata* (brides veil), and *Saponaria officianalis* (soap root). Saponin adjuvant formulations include purified formulations, such as QS21, as well as lipid formulations, such as ISCOMs. Saponin compositions have been purified using High Performance Thin Layer Chromatography (HP-LC) and Reversed Phase High Performance Liquid Chromatography (RP-HPLC). Specific purified fractions using these techniques have been identified, including QS7, QS17, QS18, QS21, QH-A, QH-B and QH-C. Preferably, the saponin is QS21. A method of production of QS21 is disclosed in U.S. Patent No. 5,057,540. Saponin formulations may also comprise a sterol, such as cholesterol (see WO 96/33739).

Combinations of saponins and cholesterols can be used to form unique particles called Immunostimulating Complexs (ISCOMs). ISCOMs typically also include a phospholipid such as phosphatidylethanolamine or phosphatidylcholine. Any known saponin can be used in ISCOMs. Preferably, the ISCOM includes one or more of Quil A, QHA and QHC. ISCOMs are further described in EP 0 109 942, WO 96/11711 and WO 96/33739. Optionally, the ISCOMS may be devoid of additional detergent. See WO00/07621.

A review of the development of saponin based adjuvants can be found at Barr, et al., Advanced Drug Delivery Reviews (1998) 32:247 – 271. See also Sjolander, et al., Advanced Drug Delivery Reviews (1998) 32:321 – 338.

C. Virosomes and Virus Like Particles (VLPs)

Virosomes and Virus Like Particles (VLPs) can also be used as adjuvants in the invention. These structures generally contain one or more proteins from a virus optionally combined or formulated with a phospholipid. They are generally non-pathogenic, non-replicating and generally do not contain any of the native viral genome. The viral proteins may be recombinantly produced or isolated from whole viruses. These viral proteins suitable for use in virosomes or VLPs include proteins derived from influenza virus (such as HA or NA), Hepatitis B virus (such as core or capsid proteins), Hepatitis E virus, measles virus, Sindbis virus, Rotavirus, Foot-and-Mouth Disease virus, Retrovirus, Norwalk virus, human Papilloma virus, HIV, RNA-phages, Qß-phage (such as coat proteins), GA-phage, fr-phage, AP205 phage, and Ty (such as retrotransposon Ty protein p1). VLPs are discussed further in WO 03/024480, WO 03/024481, and Niikura et al., Virology (2002) 293:273 – 280, Lenz et al., Journal of Immunology (2001) 5246 – 5355; Pinto, et al., Journal of Infectious Diseases (2003) 188:327 – 338 and Gerber et al., Journal of Virology (2001) 75(10):4752 – 4760. Virosomes are discussed further in, for example, Gluck et al., Vaccine (2002) 20:B10 –B16.

D. Bacterial or Microbial Derivatives

Adjuvants suitable for use in the invention include bacterial or microbial derivatives such as:

(1) *Non-toxic derivatives of enterobacterial lipopolysaccharide (LPS)*

Such derivatives include Monophosphoryl lipid A (MPL) and 3-O-deacylated MPL (3dMPL). 3dMPL is a mixture of 3 De-O-acylated monophosphoryl lipid A with 4, 5 or 6 acylated chains. A preferred "small particle" form of 3 De-O-acylated monophosphoryl lipid A is disclosed in EP 0 689 454. Such "small particles" of 3dMPL are small enough to be sterile filtered through a 0.22 micron membrane (see EP 0 689 454). Other non-toxic LPS derivatives include monophosphoryl lipid A mimics, such as aminoalkyl glucosaminide phosphate derivatives *e.g.* RC-529. See Johnson *et al.* (1999) *Bioorg Med Chem Lett* 9:2273-2278.

(2) *Lipid A Derivatives*

Lipid A derivatives include derivatives of lipid A from *Escherichia coli* such as OM-174. OM-174 is described for example in Meraldi et al., Vaccine (2003) 21:2485 – 2491 and Pajak, et al., Vaccine (2003) 21:836 – 842.

(3) *Immunostimulatory oligonucleotides*

Immunostimulatory oligonucleotides suitable for use as adjuvants in the invention include nucleotide sequences containing a CpG motif (a sequence containing an unmethylated cytosine followed by guanosine and linked by a phosphate bond). Bacterial double stranded RNA or oligonucleotides containing palindromic or poly(dG) sequences have also been shown to be immunostimulatory.

The CpG's can include nucleotide modifications/analogs such as phosphorothioate modifications and can be double-stranded or single-stranded. Optionally, the guanosine may be replaced with an analog such as 2'-deoxy-7-deazaguanosine. See Kandimalla, et al., Nucleic Acids Research (2003) 31(9): 2393 – 2400; WO 02/26757 and WO 99/62923 for examples of possible analogue substitutions. The adjuvant effect of CpG oligonucleotides is further discussed in Krieg, Nature Medicine (2003) 9(7): 831 – 835; McCluskie,

et al., FEMS Immunology and Medical Microbiology (2002) 32:179 – 185; WO 98/40100, U.S. Patent No. 6,207,646, U.S. Patent No. 6,239,116, and U.S. Patent No. 6,429,199.

The CpG sequence may be directed to TLR9, such as the motif GTCGTT or TTCGTT. See Kandimalla, et al., Biochemical Society Transactions (2003) 31 (part 3): 654 – 658. The CpG sequence may be specific for inducing a Th1 immune response, such as a CpG-A ODN, or it may be more specific for inducing a B cell response, such a CpG-B ODN. CpG-A and CpG-B ODNs are discussed in Blackwell, et al., J. Immunol. (2003) 170(8):4061 – 4068; Krieg, TRENDS in Immunology (2002) 23(2): 64 – 65 and WO 01/95935. Preferably, the CpG is a CpG-A ODN.

Preferably, the CpG oligonucleotide is constructed so that the 5' end is accessible for receptor recognition. Optionally, two CpG oligonucleotide sequences may be attached at their 3' ends to form "immunomers". See, for example, Kandimalla, et al., BBRC (2003) 306:948 – 953; Kandimalla, et al., Biochemical Society Transactions (2003) 31(part 3):664 – 658; Bhagat et al., BBRC (2003) 300:853 – 861 and WO 03/035836.

(4) *ADP-ribosylating toxins and detoxified derivatives thereof.*

Bacterial ADP-ribosylating toxins and detoxified derivatives thereof may be used as adjuvants in the invention. Preferably, the protein is derived from *E. coli* (i.e., E. coli heat labile enterotoxin "LT), cholera ("CT"), or pertussis ("PT"). The use of detoxified ADP-ribosylating toxins as mucosal adjuvants is described in WO 95/17211 and as parenteral adjuvants in WO 98/42375. Preferably, the adjuvant is a detoxified LT mutant such as LT-K63.

E. Human Immunomodulators

Human immunomodulators suitable for use as adjuvants in the invention include cytokines, such as interleukins (*e.g.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (*e.g.* interferon-γ), macrophage colony stimulating factor, and tumor necrosis factor.

F. Bioadhesives and Mucoadhesives

Bioadhesives and mucoadhesives may also be used as adjuvants in the invention. Suitable bioadhesives include esterified hyaluronic acid microspheres (Singh *et al.* (2001) *J. Cont. Rele.* 70:267-276) or mucoadhesives such as cross-linked derivatives of poly(acrylic acid), polyvinyl alcohol, polyvinyl pyrollidone, polysaccharides and carboxymethylcellulose. Chitosan and derivatives thereof may also be used as adjuvants in the invention. E.g., WO99/27960.

G. Microparticles

Microparticles may also be used as adjuvants in the invention. Microparticles (*i.e.* a particle of ~100nm to ~150μm in diameter, more preferably ~200nm to ~30μm in diameter, and most preferably ~500nm to ~10μm in diameter) formed from materials that are biodegradable and non-toxic (*e.g.* a poly(α-hydroxy acid), a polyhydroxybutyric acid, a polyorthoester, a polyanhydride, a polycaprolactone, *etc.*), with poly(lactide-co-glycolide) are preferred, optionally treated to have a negatively-charged surface (*e.g.* with SDS) or a positively-charged surface (*e.g.* with a cationic detergent, such as CTAB).

H. Liposomes

Examples of liposome formulations suitable for use as adjuvants are described in U.S. Patent No. 6,090,406, U.S. Patent No. 5,916,588, and EP 0 626 169.

I. Polyoxyethylene ether and Polyoxyethylene Ester Formulations

Adjuvants suitable for use in the invention include polyoxyethylene ethers and polyoxyethylene esters. WO99/52549. Such formulations further include polyoxyethylene sorbitan ester surfactants in combination with an octoxynol (WO01/21207) as well as polyoxyethylene alkyl ethers or ester surfactants in combination with at least one additional non-ionic surfactant such as an octoxynol (WO01/21152).

Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether (laureth 9), polyoxyethylene-9-steoryl ether, polyoxytheylene-8-steoryl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether.

J. Polyphosphazene (PCPP)

PCPP formulations are described, for example, in Andrianov et al., "Preparation of hydrogel microspheres by coacervation of aqueous polyphophazene solutions", Biomaterials (1998) 19(1 – 3):109 – 115 and Payne et al., "Protein Release from Polyphosphazene Matrices", Adv. Drug. Delivery Review (1998) 31(3):185 – 196.

K. Muramyl peptides

Examples of muramyl peptides suitable for use as adjuvants in the invention include N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), and N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine MTP-PE).

L. Imidazoquinolone Compounds.

Examples of imidazoquinolone compounds suitable for use adjuvants in the invention include Imiquamod and its homologues, described further in Stanley, "Imiquimod and the imidazoquinolones: mechanism of action and therapeutic potential" Clin Exp Dermatol (2002) 27(7):571 – 577 and Jones, "Resiquimod 3M", Curr Opin Investig Drugs (2003) 4(2):214 – 218.

The invention may also comprise combinations of aspects of one or more of the adjuvants identified above. For example, the following adjuvant compositions may be used in the invention:

(1) a saponin and an oil-in-water emulsion (WO99/11241);

(2) a saponin (e.g.., QS21) + a non-toxic LPS derivative (e.g., 3dMPL) (see WO 94/00153);

(3) a saponin (e.g.., QS21) + a non-toxic LPS derivative (e.g., 3dMPL) + a cholesterol;

(4) a saponin (*e.g.* QS21) + 3dMPL + IL-12 (optionally + a sterol) (WO98/57659);

(5) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions (European patent applications 0835318, 0735898 and 0761231);

(6) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-block polymer L121, and thr-MDP, either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion.

(7) Ribi™ adjuvant system (RAS), (Ribi Immunochem) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); and

(8) one or more mineral salts (such as an aluminum salt) + a non-toxic derivative of LPS (such as 3dPML).

(9) one or more mineral salts (such as an aluminum salt) + an immunostimulatory oligonucleotide (such as a nucleotide sequence including a CpG motif).

Aluminium salts and MF59 are preferred adjuvants for parenteral immunisation. Mutant bacterial toxins are preferred mucosal adjuvants.

The composition may include an antibiotic.

Further antigens

The compositions of the invention may further comprise one or more additional non-GAS antigens, including additional bacterial, viral or parasitic antigens.

In one embodiment, the GAS antigen combinations of the invention are combined with one or more additional, non-GAS antigens suitable for use in a paediatric vaccine. For example, the GAS antigen combinations may be combined with one or more antigens derived from a bacteria or virus selected from the group consisting of *N. meningitidis* (including serogroup A, B, C, W135 and/or Y), *Streptococcus pneumoniae*, *Bordetella pertussis, Moraxella catarrhalis, Tetanus, Diphtheria*, Respiratory Syncytial virus ('RSV'), polio, measles, mumps, rubella, and rotavirus.

In another embodiment, the GAS antigen combinations of the invention are combined with one or more additional, non-GAS antigens suitable for use in a vaccine designed to protect elderly or immunocomprised individuals. For example, the GAS antigen combinations may be combined with an antigen derived from the group consisting of *Enterococcus faecalis, Staphylococcus aureus, Staphylococcus epidermis, Pseudomonas aeruginosa, Legionella pneumophila, Listeria monocytogenes*, influenza, and Parainfluenza virus ('PIV').

Where a saccharide or carbohydrate antigen is used, it is preferably conjugated to a carrier protein in order to enhance immunogenicity {*e.g.* Ramsay *et al.* (2001) *Lancet* 357(9251):195-196; Lindberg (1999) *Vaccine* 17 Suppl 2:S28-36; Buttery & Moxon (2000) *J R Coll Physicians Lond* 34:163-168; Ahmad & Chapnick (1999) *Infect Dis Clin North Am* 13:113-133, vii.Goldblatt (1998) *J. Med. Microbiol.* 47:563-567; European patent 0 477 508; US Patent No. 5,306,492; WO98/42721; *Conjugate Vaccines* (eds. Cruse *et al.*) ISBN 3805549326, particularly vol. 10:48-114; Hermanson (1996) *Bioconjugate Techniques* ISBN: 0123423368 or 012342335X}. Preferred carrier proteins are bacterial toxins or toxoids, such as diphtheria or tetanus toxoids. The CRM_{197} diphtheria toxoid is particularly preferred {*Research Disclosure*, 453077 (Jan 2002)}. Other carrier polypeptides include the *N.meningitidis* outer membrane protein {EP-A-0372501}, synthetic peptides { EP-A-0378881 and EP-A-0427347}, heat shock proteins { WO93/17712 and WO94/03208}, pertussis proteins {WO98/58668 and EP-A-0471177}, protein D from *H.influenzae* {WO00/56360}, cytokines {WO91/01146}, lymphokines, hormones, growth factors, toxin A or B from *C.difficile* {WO00/61761}, iron-uptake proteins { WO01/72337}, *etc.* Where a mixture comprises capsular

saccharides from both serogroups A and C, it may be preferred that the ratio (w/w) of MenA saccharide:MenC saccharide is greater than 1 (*e.g.* 2:1, 3:1, 4:1, 5:1, 10:1 or higher). Different saccharides can be conjugated to the same or different type of carrier protein. Any suitable conjugation reaction can be used, with any suitable linker where necessary.

Toxic protein antigens may be detoxified where necessary *e.g.* detoxification of pertussis toxin by chemical and/or genetic means.

Where a diphtheria antigen is included in the composition it is preferred also to include tetanus antigen and pertussis antigens. Similarly, where a tetanus antigen is included it is preferred also to include diphtheria and pertussis antigens. Similarly, where a pertussis antigen is included it is preferred also to include diphtheria and tetanus antigens.

Antigens in the composition will typically be present at a concentration of at least 1μg/ml each. In general, the concentration of any given antigen will be sufficient to elicit an immune response against that antigen.

As an alternative to using protein antigens in the composition of the invention, nucleic acid encoding the antigen may be used {*e.g.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; Scott-Taylor & Dalgleish (2000) *Expert Opin Investig Drugs* 9:471-480; Apostolopoulos & Plebanski (2000) *Curr Opin Mol Ther* 2:441-447; Ilan (1999) *Curr Opin Mol Ther* 1:116-120Dubensky *et al.* (2000) *Mol Med* 6:723-732; Robinson & Pertmer (2000) *Adv Virus Res* 55:1-74Donnelly *et al.* (2000) *Am J Respir Crit Care Med* 162(4 Pt 2):S190-193Davis (1999) *Mt. Sinai J. Med.* 66:84-90}. Protein components of the compositions of the invention may thus be replaced by nucleic acid (preferably DNA *e.g.* in the form of a plasmid) that encodes the protein.

Definitions

The term "comprising" means "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The term "about" in relation to a numerical value *x* means, for example, $x \pm 10\%$.

References to a percentage sequence identity between two amino acid sequences means that, when aligned, that percentage of amino acids are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in section 7.7.18 of *Current Protocols in Molecular Biology* (F.M. Ausubel *et al.*, eds., 1987) Supplement 30. A preferred alignment is determined by the Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 2, BLOSUM matrix of 62. The Smith-Waterman homology search algorithm is disclosed in Smith & Waterman (1981) *Adv. Appl. Math.* 2: 482-489. Similar sequence identity methods can be used to determine sequence homology between two polynucleotide sequences.

The following example demonstrates one way of preparing recombinant GAS antigens of the invention and testing their efficacy in a murine model.

EXAMPLE 1: Preparation of recombinant GAS antigens of the invention and Demonstration of Efficacy in Murine Model.

Recombinant GAS proteins corresponding to two or more of the GAS antigens of the first antigen group are expressed as follows.

1. Cloning of GAS antigens for expression in E. coli

The selected GAS antigens were cloned in such a way to obtain two different kinds of recombinant proteins: (1) proteins having an hexa-histidine tag at the carboxy-terminus (Gas-His) and (2) proteins having the hexa-histidine tag at the carboxy-terminus and GST at the amino-terminus (Gst-Gas-His). Type (1) proteins were obtained by cloning in a pET21b+vector (available from Novagen). The type (2) proteins were obtained by cloning in a pGEX-NNH vector. This cloning strategy allowed for the GAS genomic DNA to be used to amplify the selected genes by PCR, to perform a single restriction enzyme digestion of the PCR products and to clone then simultaneously into both vectors.

(a) *Construction of pGEX-NNH expression vectors*

Two couples of complementary oligodeoxyribonucleotides are synthesised using the DNA synthesiser ABI394 (Perkin Elmer) and reagents from Cruachem (Glasgow, Scotland). Equimolar amounts of the oligo pairs (50 ng each oligo) are annealed in T4 DNA ligase buffer (New England Biolabs) for 10 min in a final volume of 50 µl and then left to cool slowly at room temperature. With the described procedure the following DNA linkers are obtained:

gexNN linker

```
      NdeI  NheI XmaI  EcoRI   NcoI        SalI     XhoI        SacI
GATCCCATATGGCTAGCCCGGGGAATTCGTCCATGGAGTGAGTCGACTGACTCGAGTGATCGAGCTC
    GGTATACCGATCGGGCCCCTTAAGCAGGTACCTCACTCAGCTGACTGAGCTCACTAGCTCGAG

      NotI
CTGAGCGGCCGCATGAA
GACTCGCCGGCGTACTTTCGA
```

gexNNH linker

```
      HindIII    NotI       XhoI       Hexa-Histidine
TCGACAAGCTTGCGGCCGCACTCGAGCATCACCATCACCATCACTGAT
    GTTCGAACGCCGGCGTGAGCACGTAGAGGTAGTGGTAGTGACTATCGA
```

The plasmid pGEX-KG [K. L. Guan and J. E. Dixon, *Anal. Biochem.* **192**, 262 (1991)] is digested with BamHI and HindIII and 100 ng is ligated overnight at 16 °C to the linker gexNN with a molar ratio of 3:1 linker/plasmid using 200 units of T4 DNA ligase (New england Biolabs). After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NN plasmid, having the correct linker, is selected by means of restriction enzyme analysis and DNA sequencing.

The new plasmid pGEX-NN is digested with SalI and HindIII and ligated to the linker gexNNH. After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NNH plasmid, having the correct linker, is selected by means of restriction enzyme analysis and DNA sequencing.

(b) *Chromosomal DNA preparation*

GAS SF370 strain is grown in THY medium until OD_{600} is 0.6-0.8. Bacteria are then centrifuged, suspended in TES buffer with lyzozyme (10mg/ml) and mutanolysine (10U/µl) and incubated 1 hr at 37° C.

Following treatment of the bacterial suspension with RNAase, Proteinase K and 10% Sarcosyl/EDTA, protein extraction with saturated phenol and phenol/chloroform is carried out. The resulting supernatant is precipitated with Sodium Acetate/Ethanol and the extracted DNA is pelletted by centrifugation, suspended in Tris buffer and kept at –20° C.

(c) *Oligonucleotide design*

Synthetic oligonucleotide primers are designed on the basis of the coding sequence of each GAS antigen using the sequence of *Streptococcus pyogenes* SF370 M1 strain. Any predicted signal peptide is omitted, by deducing the 5' end amplification primer sequence immediately downstream from the predicted leader sequence. For most GAS antigens, the 5' tail of the primers (see Table 1, below) include only one restriction enzyme recognition site (NdeI, or NheI, or SpeI depending on the gene's own restriction pattern); the 3' primer tails (see Table 1) include a XhoI or a NotI or a HindIII restriction site.

5' tails		3' tails	
NdeI	5' GTGCGT**CATATG** 3'	XhoI	5' GCGT**CTCGAG** 3'
NheI	5' GTGCGT**GCTAGC** 3'	NotI	5' ACTCGCTA**GCGGCCGC** 3'
SpeI	5' GTGCGT**ACTAGT** 3'	HindIII	5' GCGT**AAGCTT** 3'

Table 1. Oligonucleotide tails of the primers used to amplify genes encoding selected GAS antigens.

As well as containing the restriction enzyme recognition sequences, the primers include nucleotides which hybridize to the sequence to be amplified. The number of hybridizing nucleotides depends on the melting temperature of the primers which can be determined as described [(Breslauer et al., Proc. Nat. Acad. Sci. 83, 3746-50 (1986)]. The average melting temperature of the selected oligos is 50-55 °C for the hybridizing region alone and 65-75 °C for the whole oligos. Oligos can be purchased from MWG-Biotech S.p.A. (Firenze, Italy).

(d) *PCR amplification*

The standard PCR protocol is as follows: 50 ng genomic DNA are used as template in the presence of 0,2 µM each primer, 200 µM each dNTP, 1,5 mM $MgCl_2$, 1x PCR buffer minus Mg (Gibco-BRL), and 2 units of Taq DNA polymerase (Platinum Taq, Gibco-BRL) in a final volume of 100 µl. Each sample undergoes a double-step amplification: the first 5 cycles are performed using as the hybridizing temperature of one of the oligos excluding the restriction enzyme tail, followed by 25 cycles performed according to the hybridization temperature of the whole length primers. The standard cycles are as follows:

one cycle:
denaturation : 94 °C, 2 min,

5 cycles:
denaturation: 94 °C, 30 seconds,
hybridization: 51 °C, 50 seconds,
elongation: 72 °C, 1 min or 2 min and 40 sec,

25 cycles:
denaturation: 94 °C, 30 seconds,
hybridization: 70 °C, 50 seconds,
elongation: 72 °C, 1 min or 2 min and 40 sec,

72 °C, 7 min,
4 °C

The elongation time is 1 min for GAS antigens encoded by ORFs shorter than 2000 bp, and 2 min and 40 seconds for ORFs longer than 2000 bp. The amplifications are performed using a Gene Amp PCR system 9600 (Perkin Elmer).

To check the amplification results, 4 µl of each PCR product is loaded onto 1-1.5 agarose gel and the size of amplified fragments compared with DNA molecular weight standards (DNA markers III or IX, Roche). The PCR products are loaded on agarose gel and after electrophoresis the right size bands are excised from the gel. The DNA is purified from the agarose using the Gel Extraction Kit (Qiagen) following the instruction of the manufacturer. The final elution volume of the DNA is 50 µl TE (10 mM Tris-HCl, 1 mM EDTA, pH 8). One µl of each purified DNA is loaded onto agarose gel to evaluate the yield.

(e) *Digestion of PCR fragments*

One-two µg of purified PCR products are double digested overnight at 37 °C with the appropriate restriction enzymes (60 units of each enzyme) using the appropriate restriction buffer in 100 µl final volume. The restriction enzymes and the digestion buffers are from New England Biolabs. After purification of the digested DNA (PCR purification Kit, Qiagen) and elution with 30 µl TE, 1 µl is subjected to agarose gel electrophoresis to evaluate the yield in comparison to titrated molecular weight standards (DNA markers III or IX, Roche).

(f) *Digestion of the cloning vectors (pET21b+ and pGEX-NNH)*

10 µg of plasmid is double digested with 100 units of each restriction enzyme in 400 µl reaction volume in the presence of appropriate buffer by overnight incubation at 37 °C. After electrophoresis on a 1% agarose gel, the band corresponding to the digested vector is purified from the gel using the Qiagen Qiaex II Gel Extraction Kit and the DNA was eluted with 50 µl TE. The DNA concentration is evaluated by measuring OD_{260} of the sample.

(g) *Cloning of the PCR products*

Seventy five ng of the appropriately digested and purified vectors and the digested and purified fragments corresponding to each selected GAS antigen are ligated in final volumes of 10-20 µl with a molar ratio of 1:1 fragment/vector, using 400 units T4 DNA ligase (New England Biolabs) in the presence of the buffer supplied by the manufacturer. The reactions are incubated overnight at 16 °C.

Transformation of *E coli* BL21 (Novagen) and *E coli* BL21-DE3 (Novagen) electrocompetent cells is performed using pGEX-NNH ligations and pET21b+ ligations respectively. The transformation procedure is as follows: 1-2 µl the ligation reaction is mixed with 50 µl of ice cold competent cells, then the cells are poured in a gene pulser 0.1 cm electrode cuvette (Biorad). After pulsing the cells in a MicroPulser electroporator (Biorad) following the manufacturer instructions the cells are suspended in 0.95 ml of SOC medium and incubated for 45 min at 37 °C under shaking. 100 and 900 µl of cell suspensions are plated on separate plates of agar LB 100 µg/ml Ampicillin and the plates are incubated overnight at 37 °C. The screening of the transformants is done by PCR: randomly chosen transformants are picked and suspended in

30 µl of PCR reaction mix containing the PCR buffer, the 4 dNTPs, 1,5 mM $MgCl_2$, Taq polymerase and appropriate forward and reverse oligonucleotide primers that are able to hibridize upstream and downstream from the polylinker of pET21b+ or pGEX-NNH vectors. After 30 cycles of PCR, 5 µl of the resulting products are run on agarose gel electrophoresis in order to select for positive clones from which the expected PCR band is obtained. PCR positive clones are chosen on the basis of the correct size of the PCR product, as evaluated by comparison with appropriate molecular weight markers (DNA markers III or IX, Roche).

2. Protein expression

PCR positive colonies are inoculated in 3 ml LB 100 µg/ml Ampicillin and grown at 37 °C overnight. 70 µl of the overnight culture is inoculated in 2 ml LB/Amp and grown at 37 °C until OD_{600} of the pET clones reached the 0,4-0,8 value or until OD_{600} of the pGEX clones reached the 0,8-1 value. Protein expression is then induced by adding 1 mM IPTG (Isopropil β-D thio-galacto-piranoside) to the mini-cultures. After 3 hours incubation at 37 °C the final OD_{600} is checked and the cultures are cooled on ice. After centrifugation of 0.5 ml culture, the cell pellet is suspended in 50 µl of protein Loading Sample Buffer (60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% w/v Bromophenol Blue, 100 mM DTT) and incubated at 100 °C for 5 min. A volume of boiled sample corresponding to 0.1 OD_{600} culture is analysed by SDS-PAGE and Coomassie Blue staining to verify the presence of induced protein band.

3. Purification of the recombinant proteins

Single colonies are inoculated in 25 ml LB 100 µg/ml Ampicillin and grown at 37 °C overnight. The overnight culture is inoculated in 500 ml LB/Amp and grown under shaking at 25 °C until OD_{600} 0.4-0.7. Protein expression is then induced by adding 1 mM IPTG to the cultures. After 3.5 hours incubation at 25 °C the final OD_{600} is checked and the cultures are cooled on ice. After centrifugation at 6000 rpm (JA10 rotor, Beckman), the cell pellet is processed for purification or frozen at –20° C.

(a) *Procedure for the purification of soluble His-tagged proteins from E.coli*

(1) Transfer the pellets from –20°C to ice bath and reconstitute with 10 ml 50 mM $NaHPO_4$ buffer, 300 mM NaCl, pH 8,0, pass in 40-50 ml centrifugation tubes and break the cells as per the following outline.

(2) Break the pellets in the French Press performing three passages with in-line washing.

(3) Centrifuge at about 30-40000 x g per 15-20 min. If possible use rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.)

(4) Equilibrate the Poly-Prep columns with 1 ml Fast Flow Chelating Sepharose resin with 50 mM phosphate buffer, 300 mM NaCl, pH 8,0.

(5) Store the centrifugation pellet at –20°C, and load the supernatant in the columns.

(6) Collect the flow through.

(7) Wash the columns with 10 ml (2 ml + 2 ml + 4 ml) 50 mM phosphate buffer, 300 mM NaCl, pH 8.0.

(8) Wash again with 10 ml 20 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8.0.

(9) Elute the proteins bound to the columns with 4.5 ml (1.5 ml + 1.5 ml + 1.5 ml) 250 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8,0 and collect the 3 corresponding fractions of ~1.5 ml each. Add to each tube 15 µl DTT 200 mM (final concentration 2 mM)

(10) Measure the protein concentration of the first two fractions with the Bradford method, collect a 10 μg aliquot of proteins from each sample and analyse by SDS-PAGE. (N.B.: should the sample be too diluted, load 21 μl + 7 μl loading buffer).

(11) Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.

(12) For immunisation prepare 4-5 aliquots of 100 μg each in 0.5 ml in 40% glycerol. The dilution buffer is the above elution buffer, plus 2 mM DTT. Store the aliquots at –20°C until immunisation.

(b) *Purification of His-tagged proteins from Inclusion bodies*

Purifications are carried out essentially according the following protocol:

(1) Bacteria are collected from 500 ml cultures by centrifugation. If required store bacterial pellets at -20°C. For extraction, resuspend each bacterial pellet in 10 ml 50 mM TRIS-HCl buffer, pH 8,5 on an ice bath.

(2) Disrupt the resuspended bacteria with a French Press, performing two passages.

(3) Centrifuge at 35000 x g for 15 min and collect the pellets. Use a Beckman rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.).

(4) Dissolve the centrifugation pellets with 50 mM TRIS-HCl, 1 mM TCEP {Tris(2-carboxyethyl)-phosphine hydrochloride, Pierce} , 6M guanidium chloride, pH 8.5. Stir for ~ 10 min. with a magnetic bar.

(5) Centrifuge as described above, and collect the supernatant.

(6) Prepare an adequate number of Poly-Prep (Bio-Rad) columns containing 1 ml of Fast Flow Chelating Sepharose (Pharmacia) saturated with Nichel according to manufacturer recommendations.. Wash the columns twice with 5 ml of H_20 and equilibrate with 50 mM TRIS-HCl, 1 mM TCEP, 6M guanidinium chloride, pH 8.5.

(7) Load the supernatants from step 5 onto the columns, and wash with 5 ml of 50 mM TRIS-Hcl buffer, 1 mM TCEP, 6M urea, pH 8.5

(8) Wash the columns with 10 ml of 20 mM imidazole, 50 mM TRIS-HCl , 6M urea, 1 mM TCEP, pH 8.5. Collect and set aside the first 5 ml for possible further controls.

(9) Elute the proteins bound to the columns with 4.5 ml of a buffer containing 250 mM imidazole, 50 mM TRIS-HCl, 6M urea, 1 mM TCEP, pH 8.5. Add the elution buffer in three 1.5 ml aliquots, and collect the corresponding 3 fractions. Add to each fraction 15 μl DTT (final concentration 2 mM).

(10) Measure eluted protein concentration with the Bradford method, and analyse aliquots of ca 10 μg of protein by SDS-PAGE.

(11) Store proteins at -20°C in 40% (v/v) glycerol, 50 mM TRIS-HCl, 2M urea, 0.5 M arginine, 2 mM DTT, 0.3 mM TCEP, 83.3 mM imidazole, pH 8.5.

(c) *Procedure for the purification of GST-fusion proteins from E.coli*

(1) Transfer the bacterial pellets from –20°C to an ice bath and suspend with 7,5 ml PBS, pH 7,4 to which a mixture of protease inhibitors (CØMPLETE™ - Boehringer Mannheim, 1 tablet every 25 ml of buffer) has been added.

(2) Transfer to 40-50 ml centrifugation tubes and sonicate according to the following procedure:

a. Position the probe at about 0,5 cm from the bottom of the tube
b. Block the tube with the clamp

c. Dip the tube in an ice bath

d. Set the sonicator as follows: Timer → Hold, Duty Cycle → 55, Out. Control → 6.

e. perform 5 cycles of 10 impulses at a time lapse of 1 minute (i.e. one cycle = 10 impulses + ~45" hold; b. 10 impulses + ~45" hold; c. 10 impulses + ~45" hold; d. 10 impulses + ~45" hold; e. 10 impulses + ~45" hold).

(3) Centrifuge at about 30-40000 x g for 15-20 min. E.g.: use rotor Beckman JA 25.50 at 21000 rpm, for 15 min.

(4) Store the centrifugation pellets at –20°C, and load the supernatants on the chromatography columns, as follows

(5) Equilibrate the Poly-Prep (Bio-Rad) columns with 0,5 ml (≅1 ml suspension) of Glutathione-Sepharose 4B resin, wash with 2 ml (1 + 1) H_2O, and then with 10 ml (2 + 4 + 4) PBS, pH 7,4.

(6) Load the supernatants on the columns and discard the flow through.

(7) Wash the columns with 10 ml (2 + 4 + 4) PBS, pH 7.4.

(8) Elute the proteins bound to the columns with 4.5 ml of 50 mM TRIS buffer, 10 mM reduced glutathione, pH 8.0, adding 1.5 ml + 1.5 ml + 1.5 ml and collecting the respective 3 fractions of ~1.5 ml each.

(9) Measure the protein concentration of the first two fractions with the Bradford method, analyse a 10 μg aliquot of proteins from each sample by SDS-PAGE. (N.B.: if the sample is too diluted load 21 μl (+ 7 μl loading buffer).

(10) Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.

(11) For each protein destined to the immunisation prepare 4-5 aliquots of 100 μg each in 0.5 ml of 40% glycerol. The dilution buffer is 50 mM TRIS.HCl, 2 mM DTT, pH 8.0. Store the aliquots at –20°C until immunisation.

4. Murine Model of Protection from GAS Infection

(a) *Immunization protocol*

Groups of 10 CD1 female mice aged between 6 and 7 weeks are immunized with two or more GAS antigens of the invention, (20 μg of each recombinant GAS antigen), suspended in 100 μl of suitable solution. Each group receives 3 doses at days 0, 21 and 45. Immunization is performed through intra-peritoneal injection of the protein with an equal volume of Complete Freund's Adjuvant (CFA) for the first dose and Incomplete Freund's Adjuvant (IFA) for the following two doses. In each immunization scheme negative and positive control groups are used.

For the negative control group, mice are immunized with *E. coli* proteins eluted from the purification columns following processing of total bacterial extract from a E. coli strain containing either the pET21b or the pGEX-NNH vector (thus expressing GST only) without any cloned GAS ORF (groups can be indicated as HisStop or GSTStop respectively).

For the positive control groups, mice are immunized with purified GAS M cloned from either GAS SF370 or GAS DSM 2071 strains (groups indicated as 192SF and 192DSM respectively).

Pooled sera from each group is collected before the first immunization and two weeks after the last one. Mice are infected with GAS about a week after.

Immunized mice are infected using a GAS strain different from that used for the cloning of the selected proteins. For example, the GAS strain can be DSM 2071 M23 type, obtainable from the German Collection of Microorganisms and Cell Cultures (DSMZ).

For infection experiments, DSM 2071 is grown at 37° C in THY broth until OD_{600} 0.4. Bacteria are pelletted by centrifugation, washed once with PBS, suspended and diluted with PBS to obtain the appropriate concentration of bacteria/ml and administered to mice by intraperitoneal injection. Between 50 and 100 bacteria are given to each mouse, as determined by plating aliquots of the bacterial suspension on 5 THY plates. Animals are observed daily and checked for survival.

5. Analysis of Immune Sera

(a) *Preparation of GAS total protein extracts*

Total protein extracts are prepared by incubating a bacterial culture grown to OD_{600} 0.4-0.5 in Tris 50mM pH 6.8/mutanolysin (20 units/ml) for 2 hr at 37° C, followed by incubation for ten minutes on ice in 0.24 N NaOH and 0.96% β-mercaptoethanol. The extracted proteins are precipitated by addition of trichloroaceticacid, washed with ice-cold acetone and suspended in protein loading buffer.

(b) *Western blot analysis*

Aliquots of total protein extract mixed with SDS loading buffer (1x: 60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% Bromophenol Blue, 100 mM DTT) and boiled 5 minutes at 95° C, were loaded on a 12.5% SDS-PAGE precast gel (Biorad). The gel is run using a SDS-PAGE running buffer containing 250 mM TRIS, 2.5 mM Glycine and 0.1 %SDS. The gel is electroblotted onto nitrocellulose membrane at 200 mA for 60 minutes. The membrane is blocked for 60 minutes with PBS/0.05 % Tween-20 (Sigma), 10% skimmed milk powder and incubated O/N at 4° C with PBS/0.05 % Tween 20, 1% skimmed milk powder, with the appropriate dilution of the sera. After washing twice with PBS/0.05 % Tween, the membrane is incubated for 2 hours with peroxidase-conjugated secondary anti-mouse antibody (Amersham) diluted 1:4000. The nitrocellulose is washed three times for 10 minutes with PBS/0.05 % Tween and once with PBS and thereafter developed by Opti-4CN Substrate Kit (Biorad).

(c) *Preparation of Paraformaldehyde treated GAS cultures*

A bacterial culture grown to OD_{600} 0.4-0.5 is washed once with PBS and concentrated four times in PBS/0.05 % Paraformaldehyde. Following 1 hr incubation at 37° C with shacking, the treated culture is kept overnight at 4° C and complete inactivation of bacteria is then controlled by plating aliquots on THY blood agar plates.

(d) *FACS analysis of Paraformaldehyde treated GAS coltures with mouse immune sera*

About 10^5 Paraformaldehyde inactivated bacteria are washed with 200 µl of PBS in a 96 wells U bottom plate and centrifuged for 10 min. at 3000g, at 4°C. The supernatant is discarded and the bacteria are suspended in 20 µl of PBS-0.1%BSA. Eighty µl of either pre-immune or immune mouse sera diluted in PBS-0.1%BSA are added to the bacterial suspension to a final dilution of either 1:100, 1:250 or 1:500, and incubated on ice for 30 min. Bacteria are washed once by adding 100 µl of PBS-0.1%BSA, centrifuged for 10 min. at 3000g, 4°C, suspended in 200 µl of PBS-0.1%BSA, centrifuged again and suspended in 10 µl of Goat Anti-Mouse IgG, $F(ab')_2$ fragment specific-R-Phycoerythrin-conjugated (Jackson Immunoresearch

Laboratories Inc., cat.N°115-116-072) in PBS-0.1%BSA to a final dilution of 1:100, and incubated on ice for 30 min. in the dark. Bacteria are washed once by adding 180 µl of PBS-0.1%BSA and centrifuged for 10 min. at 3000g, 4°C. The supernatant is discarded and the bacteria were suspended in 200 µl of PBS. Bacterial suspension is passed through a cytometric chamber of a FACS Calibur (Becton Dikinson, Mountain View, CA USA) and 10.000 events are acquired. Data are analysed using Cell Quest Software (Becton Dikinson, Mountain View, CA USA) by drawing a morphological dot plot (using forward and side scatter parameters) on bacterial signals. An histogram plot is then created on FL2 intensity of fluorescence log scale recalling the morphological region of bacteria.

EXAMPLE 2: Comparison of virulence of wild type GAS strain (including GAS 40) and GAS 40 deletion mutant.

The following example provides a comparison between the virulence of a wild type GAS strain and a GAS 40 deletion mutant. Mutant GAS strains where a majority of the GAS 40 sequence is removed were prepared by standard methods. Immunization groups of ten mice per group were injected with either the wild type or mutant GAS strains. As shown below, injection of a range of concentrations of the wild type isolate resulted in mouse fatalities, while injection with the GAS Δ40 mutant did not.

GAS strain	concentration	number of fatalities
wild type	2×10^5	10
wild type	2×10^6	9
wild type	2×10^7	5
GAS Δ40	2×10^2	0
GAS Δ40	2×10^3	0
GAS Δ40	2×10^4	0
GAS Δ40	2×10^5	0
GAS Δ40	2×10^6	0
GAS Δ40	2×10^7	0

EXAMPLE 3: Bacterial Opsonophagocytosis assay of GAS 40 constructs

The following example demonstrates the surface exposure of GAS 40 by use in a bacterial opsonophagocytosis assay. The following GAS constructs, each of which is described in detail above, were used in the assay: 40a-CH, 40a-RR-NH, 40a-RR, GST-40, 40a, 40a and 40a-NH. (The two references to "40a" in Figure 7 refer to sera prepared on different days.

The assay was performed as follows.

1. Preparation of bacterial inoculum. GAS bactera are grown in THY medium until they reach the middle exponential phase (OD_{600} 0.4) at 37°C. Bacteria are washed twice in chilled saline solution and are suspended in HBSS medium with the volume being adjusted for each strain depending on the amount of bacteria which will be used. Bacterial cells are kept in ice until use.

2. Preparation of PMN. PMN are prepared from buffy coats of heparinized blood from healty volunteers. The buffy coat is incubated for 30 minutes in a solution containing dextran, NaCl and

Heparin (rate 1:1). After incubation the supernatant, rich of leukocytes, is removed, transferred in a clean tube and centrifuged at 700xg for 20 minutes. A short wash in water is performed to break red blood cells and then a solution of NaCl is added to restore the appropriate salt concentration. After this step cells are centrifuged, washed and suspended in MEM at a suitable concentration.

3. Opsonophagocytosis assay. GAS strains (prepared as described) are incubated with heat inactivated immune mice serum derived from immunization with the indicated GAS antigen (or preimmune for the control), human PMN and baby rabbit complement. 1 hour of incubation at 37°C. Samples taken immediately before and after the incubation are plated on THY blood agar plates. Phagocytosis is evaluated comparing the difference in the number of colonies at the two times for the preimmune and the immune serum. Data are reported as logarithm number of grown colonies at t=0 - logarithm number of grown colonies at t=60

The results of the assay are shown in Figure 7. The Y axis reports the difference between the logarithm of colony counts at time 0 and the logarithm of the colony counts after 60 seconds: log(CFU @ T0)-log(CFU @ T60'). If there has been growth (*i.e.*, the bacteria are not activelly killed), negative numbers (negative bars) result. If bacteria are killed, positive numbers (positive histogram bars) result. As shown in Figure 7, positive histogram bars are reported for each of the GAS constructs. The last four yellow bars in Figure 7 represent controls: **B**= bacteria alone, **B PMN**= bacteria + polymorphonucleates, **B C**= Bacteria + complement, **P PMN C**= bacteria + polymorphonucleates + complement (no serum).

EXAMPLE 4: GAS 40 immunization challenge experiments in murine mouse model of protection

A sample of the percent survival results from numerous murine mouse model experiments using the GAS 40 antigen are listed below. Annotations indicate where construct used to express the recombinant GAS 40 antigen was modified to facilitate expression.

GAS antigen	% Survival in Mouse Challenge Model
40a	55
40a-RR	70
40a-RR-NH	60

It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention.

CLAIMS:

1. An immunogenic composition comprising a combination of GAS antigens, said combination consisting of two to ten GAS antigens, wherein said combination includes GAS 40.

2. The composition of claim 1, wherein said combination of GAS antigens further includes one or more GAS antigens selected from the group consisting of GAS 39, GAS 57, GAS 117, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511.

3. The composition of claim 1, wherein said combination of GAS antigens includes GAS 117.

4. The composition of claim 1, wherein said GAS 40 antigen comprises an amino acid sequence comprising a first coiled-coil region and a second coiled-coil region.

5. The composition of claim 1, wherein the GAS 40 antigen comprises an amino acid sequence comprising a first coiled-coil region.

6. The composition of claim 5, wherein said first coiled-coil region comprises an amino acid sequence comprising SEQ ID NO: 12.

7. The composition of claim 4, wherein the GAS 40 antigen comprises an amino acid sequence comprising a second coiled-coil region.

8. The composition of claim 7, wherein said second coiled-coil region includes a leucine zipper region.

9. The composition of claim 7, wherein the second coiled-coil region comprises an amino acid sequence comprising SEQ ID NO: 13.

10. An immunogenic composition comprising a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of a first antigen group, said first antigen group consisting of GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057.

11. The immunogenic composition of claim 10, wherein said combination of GAS antigens is selected from the group consisting of GAS 39, GAS 40, GAS 57, GAS 117, GAS 202, GAS 294, GAS 527, GAS 533, and GAS 511.

12. The immunogenic composition of claim 10, wherein said combination of GAS antigens includes GAS 40 and GAS 117.

13. The immunogenic composition of claim 10, wherein said combination includes GAS 40.

14. The immunogenic composition of claim 10, wherein said GAS 40 is selected from the amino acid sequence comprising (a) a first coiled-coil region, (b) a second coiled-coil region or (c) a first coiled-coil region and a second coiled-coil region.

15. A fusion construct comprising a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of a first antigen group, said first antigen group consisting of GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057.

16. The fusion construct of claim 15, wherein said combination includes GAS 40.

17. The fusion construct of claim 15, wherein said combination includes GAS 40 and GAS 117.

18. A composition comprising a combination of two or more antibodies selected from the group consisting of antibodies specific to antibodies comprising an antibody specific to GAS 40, GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057.

19. The composition of claim 18, wherein said combination includes an antibody specific to GAS 40.

20. The composition of claim 19, wherein said GAS 40 specific antibody is specific to the first or the second coiled-coil region of GAS 40.

21. A method for the therapeutic or prophylactic treatment of Streptococcus pyogenes infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of any one of the immunogenic compositions of claims 1 to 20.

22. A method of manufacturing any one of the immunogenic compositions of claims 1 to 20.

23. A kit comprising a first component comprising any one of the immunogenic compositions of claims 1 to 20.

24. A composition comprising a GAS 40 antigen, wherein said antigen comprises an amino acid sequence comprising a first coiled-coil region or a second coiled-coil region.

25. The composition of claim 24, wherein said GAS 40 antigen comprises a first coiled-coil region comprising SEQ ID NO: 12.

26. The composition of claim 24, wherein said GAS 40 antigen comprises a second coiled-coil region comprising SEQ ID NO: 13.

27. An antibody specific to any one of the compositions of claims 24 to 26.

FIGURE 1: Annotation of GAS 40

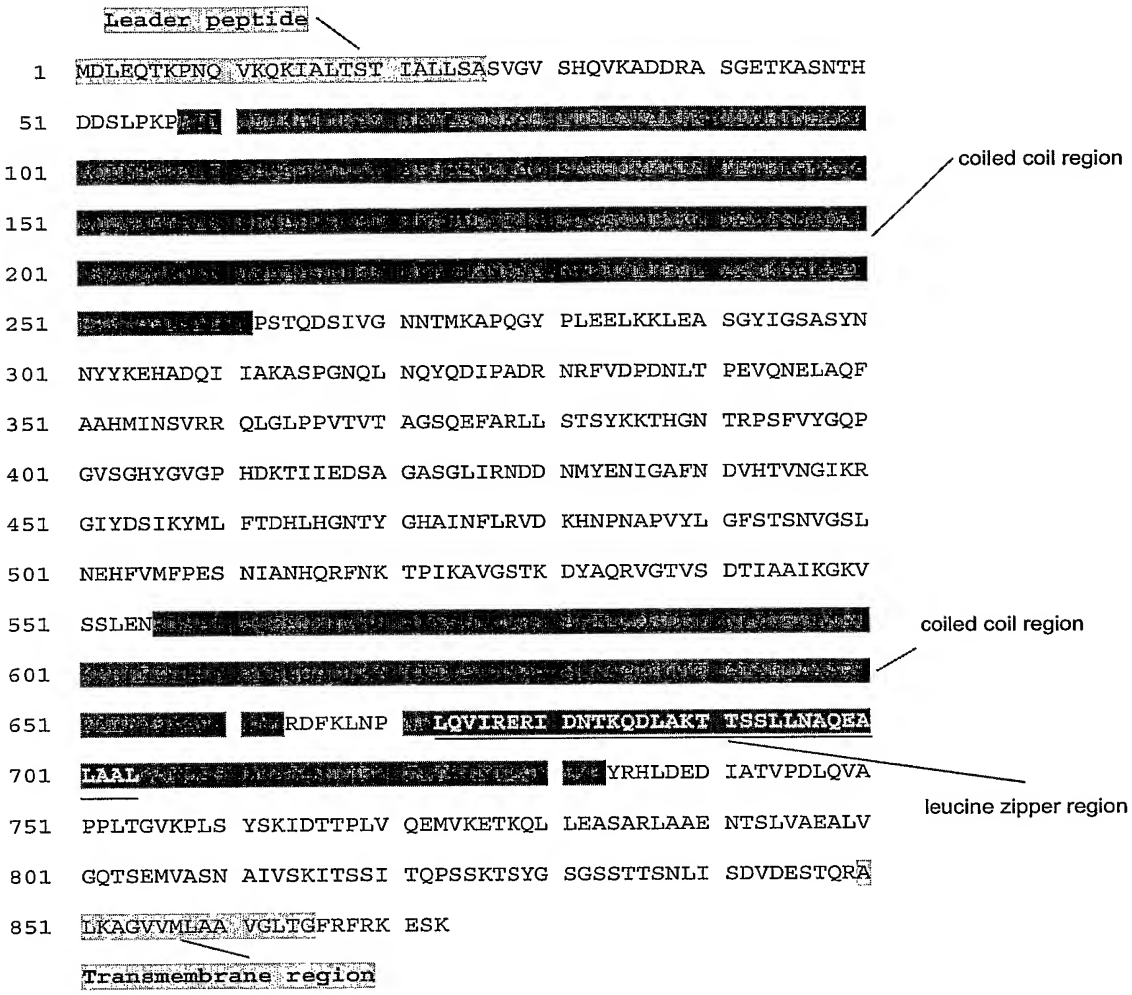
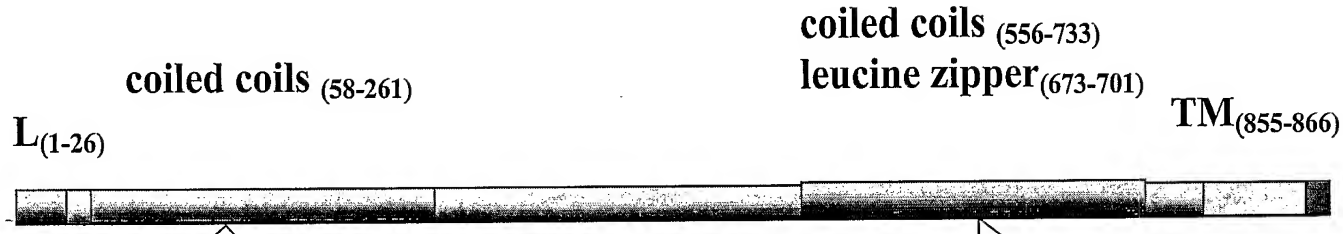


FIGURE 2 : Schematic of GAS40: putative surface exclusion protein prgA (873aa)



(aa 33-324) 22% identity to streptococcal surface protein A precursor from *Streptococcus gordonii.*
surface protein adhesin; member of streptococcal antigen I/II family of protein adhesins recognizing salivary agglutinin glycoprotein (gp-340), collagen type I

(aa 509-717) 46% identity to the immunoreactive protein Se89.9 (210 aa fragment) from *Streptococcus equi.*
Identification and Genomic Organization of Genes for Immunoreactive Surface Exposed Proteins of *Streptococcus equi*

(aa 43-238) 23% identity to surface protein pspA precursor from *Streptococcus pneumoniae.*
This aminoterminal domain of this *S. pn* protein is essential for full pneumococcal virulence, and monoclonal antibodies raised against it protect mice against pneumococcal infections. This domain has a monomeric form with an axial shape ratio of approximately 1:12, typical of fibrous proteins. Sequence analyses indicated an alpha-helical coiled-coil structure for this monomeric molecule with only few loop-type breaks in helicity.

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other Streptococcus bacteria

3(a) BLAST alignment of amino acid sequence of GAS 40 including the first coiled-coil region with SpA precursor of *Streptococcus gordonii*

```
>gi|25990270|gb|AAC44101.3| streptococcal surface protein A precursor
[Streptococcus gordonii]
          Length = 1575
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus pyogenes]
          Length = 873

 Score = 63.2 bits (152), Expect = 5e-11
 Identities = 65/293 (22%), Positives = 124/293 (42%), Gaps = 13/293
(4%)

Query: 112 QDQTSDKGTATTAAENAQKQAEIKSDYAKQA----EEIKKTTEAYKKEVEAHQAETDKIN
167
           Q +  D+ +  T A N    +  K +  ++A    + ++KT    K E+         K
Sbjct: 33  QVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVEKTLSQQKAELTELATALTKTT
92

Query: 168 AENKAAEDKYQEDLKAHQAEVEKINTANATAKAEYEAKLAQYQKDLAAVQKANEDSQLDY
227
           AE    +++   + KA  +  E       A+++    A+ A++Q++L A +    ++Q D
Sbjct: 93  AEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATETELHNAQADQ
152

Query: 228 QNKLSAYQAELARVQKANAEAKEAYE--KAVKENTAKNAALQAENEAIKQRNETAKANYD
285
            +K +A   + A +      A++  E  K  ++N AK  A+ +  +AI +  +TA  N
Sbjct: 153 HSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTK
212

Query: 286 AAMKQYEADLAAIKKAKEDNDADYQAKLAAYQAELARVQKANADAKAAYEKAVEENTAKN
345
           A   + E   A ++  K          +LAA +A LA  +   +  K++     +     N
Sbjct: 213 ALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSIVGNN
272

Query: 346 TAIQAEN---EAIKQRNAA---AKATYEAALKQYEADLAAAKKANEDSDADYQ 392
           T    +    E +K+  A+      A+Y    K++ AD   AK +  +    YQ
Sbjct: 273 TMKAPQGYPLEELKKLEASGYIGSASYNNYYKEH-ADQIIAKASPGNQLNQYQ 324
```

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other Streptococcus bacteria

3(b) BLAST alignment of amino acid sequence of GAS 40 including the first coiled-coil region with SpB precursor of *Streptococcus gordonii*

```
>gi|25055226|gb|AAC44102.3|    streptococcal surface protein B precursor
[Streptococcus gordonii]
          Length = 1499
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
          Length = 873

 Score = 54.3 bits (129), Expect = 2e-08
 Identities = 53/226 (23%), Positives = 98/226 (43%), Gaps = 13/226
(5%)

Query: 111 QDQTSDKGTATTAAENAQKQAEIKSDYAKQA----EEIKKTTEAYKKEVEAHQAETDKIN
166
           Q +  D+ +  T A N    +  K +  ++A    + ++KT    K E+         K
Sbjct: 33  QVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVEKTLSQQKAELTELATALTKTT
92

Query: 167 AENKAAEDKYQEDLKAHQAEVEKINTANATAKAEYEAKLAQYQKDLAAVQKANEDSQLDY
226
           AE    +++   + KA  +  E      A+++    A+ A++Q++L A +    ++Q D
Sbjct: 93  AEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATETELHNAQADQ
152

Query: 227 QNKLSAYQAELARV--QXXXXXXXXXXXXXXXXNTAKNAALQAENEAIKQRNETAKANYD
284
            +K +A   + A +  +                  N AK  A+ +  +AI +  +TA  N
Sbjct: 153 HSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTK
212

Query: 285 AAMKQYE---ADL----AAIKKAKEDNDADYQAKLAAYQAELARVQ 323
           A   + E   ADL    A +KK   +  A  +A LA  +AEL+R++
Sbjct: 213 ALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLK 258
```

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other Streptococcus bacteria

3(c) BLAST alignment of amino acid sequence of GAS 40 including the first coiled-coil region with Surface Protein PspA precursor of *Streptococcus pneumoniae*

```
>gi| 282335 |pir||A41971 surface protein pspA precursor - Streptococcus
pneumoniae
>ref|NP 268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
          Length = 873

 Score = 48.1 bits (113), Expect = 6e-07
 Identities = 46/200 (23%), Positives = 89/200 (44%), Gaps = 23/200
(11%)

Query: 139 KTKFNTVRAMVVPEPEQLAETK-------KKSEEAKQKAPELTKKLEEAKAKLEE-AEKK
190
           +TK +      +P+PE + E K       K   + K +  EL   L +  A++    E++
Sbjct: 43  ETKASNTHDDSLPKPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQ
102

Query: 191 ATEAKQKVDAEEVAPQAKIAELENQVHRLEQELKEIDESESEDYAKEGFRAPLQSKLDAK
250
             E K    A+E+      +  E  + +  +  +E+  +E+E +  +  +   ++ L  +
Sbjct: 103 DNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATETELHNAQADQHSKETALSEQ
162

Query: 251 KAKLS----KLEELSDKIDELDAEIAKLEDQL-------KAAEENNNVEDYFKEGLEKTI
299
           KA +S    + ++L +++    +  IAKL   +       KAA+  N+        LEK
Sbjct: 163 KASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTKALSSELEKA-
221

Query: 300 AAKKAELEKTEADLKKAVNE 319
               KA+LE  +A +KK + E
Sbjct: 222 ---KADLENQKAKVKKQLTE 238
```

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other Streptococcus bacteria

3(d) BLAST alignment of amino acid sequence of GAS 40 including the second coiled-coil region with SpB precursor of *Streptococcus gordonii*

```
>gi|23380384|gb|AAN18299.1|    immunoreactive protein Se89.9 (fragment)
[Streptococcus equi]
      Length = 210
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
          Length = 873

 Score =  173 bits (438), Expect = 4e-45
 Identities = 98/209 (46%), Positives = 144/209 (68%)

Query: 1   ESDIVDATRFSTTEIPKSGQVIDRSASIQALTNDIASIKGKIASLESRLADPSSEAEVTA
60
           ES+I +  RF+ T I   G   D +  +  +++ IA+IKGK++SLE+RL+    EA++ A
Sbjct: 509 ESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGKVSSLENRLSAIHQEADIMA
568

Query: 61  AQAKISQLQHQLEAAQAKSHKLDQQVEQLANTKDSLRTQLLAAKEEQAQLKANLDKALAL
120
           AQAK+SQLQ +L +   +S  L+ QV QL +TK SLRT+LLAAK +QAQL+A  D++LA
Sbjct: 569 AQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAK
628

Query: 121 LASSKATLHKLEAAMEEAKARVAGLASQKAQLEDLLAFEKNPNRIELAQEKVAAAKKALA
180
           LAS KA LH+ EA  E+A ARV  L ++KA L+ L  F+ NPNR+++ +E++   K+ LA
Sbjct: 629 LASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLA
688

Query: 181 DTEDKLLAAQASLSDLQAQRARLQLSIAT 209
            T   LL AQ +L+ LQA+++ L+ +IAT
Sbjct: 689 KTTSSLLNAQEALAALQAKQSSLEATIAT 717
```

Figure 4: Secondary Structure Prediction of GAS 40

Figure 4(a) Secondary Structure prediction alignment with GAS 40 amino acid sequence

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MDLEQTKPNQVKQKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAV
CCCCCCCCchhhHHhhhhhHHHHhhhcccceeEEEecCCcCCCCCcCCCCCCCCCCCcHHHHHHHHHHHH
EKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTA
HHHHHHHHHHHHHHHHHHHHHhhHHHHHHHHHhhhHHHHHHHHHHHHHHhcccchHHHHHHHHHHHHHHH
TETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDN
HHHHHHHHHHcccchhHHHHhhhhcceehhhhHHHHHHHHHhhHHHHHHHHHhhhcCcHHHHHHHHHHhhc
TKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSIVGNNTMKAPQGY
cHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhcCCCCCCCceEcCCCCCCCCCC
PLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPDNLTPEVQNELAQF
CHHHHHHHhcCcccceeccchHHHHHHHHHHHHHhCCchhhhhhccCccccCCCCCCCCChHHHHHHHHH
AAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGPHDKTIIEDSA
HHHHHHHHHHHcCCCCceecCCCHHHHHHHHhhccccccCCCCceEEEcCCCeeeeeceCcCCCeEEEEcC
GASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLRVDKHNPNAPVYL
CCCceecCCcHHHhhhccccccccccCccccchHHHHHHHhheeccccCcccchhHHheeeeeecCCCCCCcEEE
GFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGKVSSLENRLSAI
EEEecCccCccccceeccccccchHHhhhCCCCCcccCCcHHHHHHHchhHHHHHHHHcCcccHHHHHHHHH
HQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLA
HHHHHHHHHHHHHHHHHHhHHHHhhccCCchhHHhhhhhcCcCHHHHHHHHHHHHHHHHHHhHHHHHHHHH
SLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEA
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHcCCccHHHHHHHHHHhHHHHHHHHHHHHHHHH
LAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATVPDLQVAPPLTGVKPLSYSKIDTTPLV
HHHHHHhhcCceeeccchHHHHHHHHHHHHhhhhhhHHHhhhccCCCccCCCCCcCCCceecCCCHHH
QEMVKETKQLLEASARLAAENTSLVAEALVGQTSEMVASNAIVSKITSSITQPSSKTSYGSGSSTTSNLI
HHHHHHHHHHHHHHHHHHHHHHhHHHHHHHHHhcchhHHHHhhchhhhcceEEecCCCccccccCccccCcce
SDVDESTQRALKAGVVMLAAVGLTGFRFRKESK
cCCchHHHHHHHhcceeeEeeccccceeeccCC
```

```
Sequence length :   873
PHD :
   Alpha helix      (Hh) :   525 is  60.14%
   3_10 helix       (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    63 is   7.22%
   Beta turn        (Tt) :     0 is   0.00%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :   285 is  32.65%
   Ambigous states  (?)  :     0 is   0.00%
```

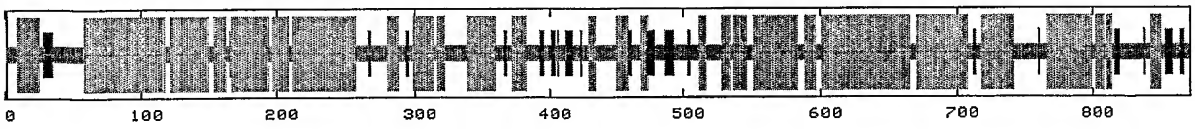
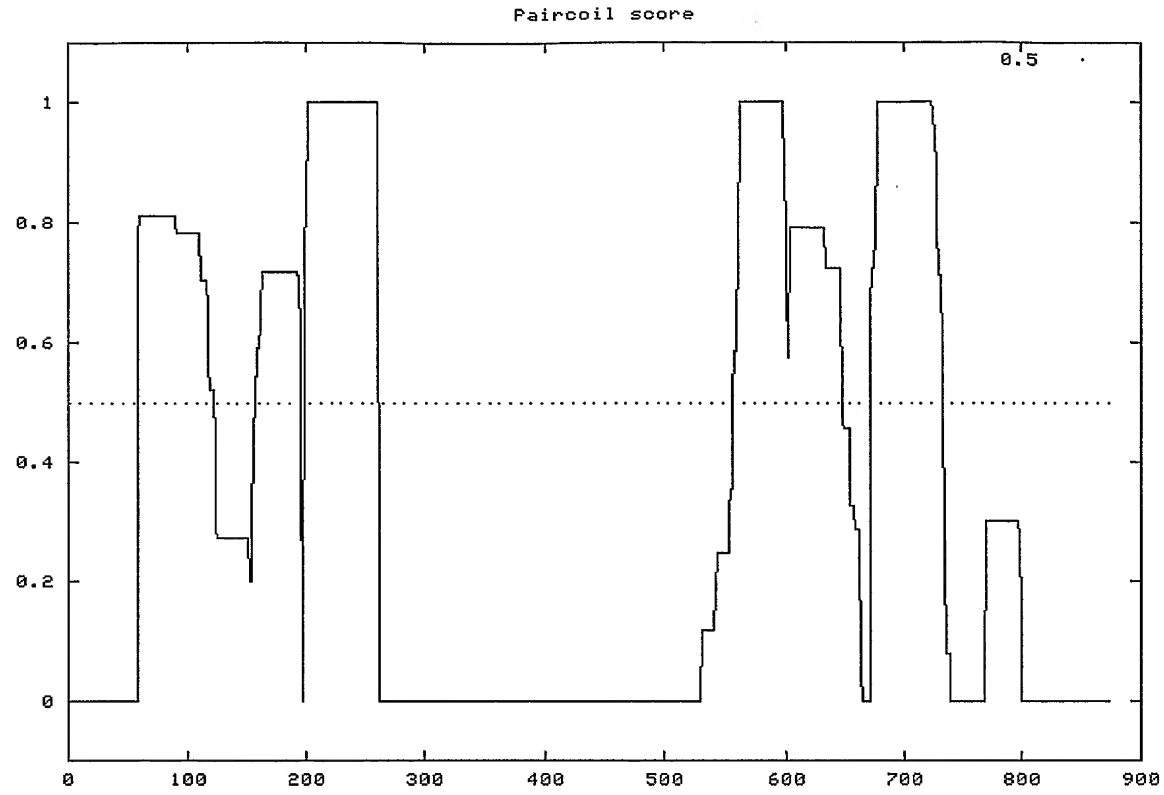


Figure 4(b): Secondary Structure prediction based on PairCoil Score



Coiled coils found:

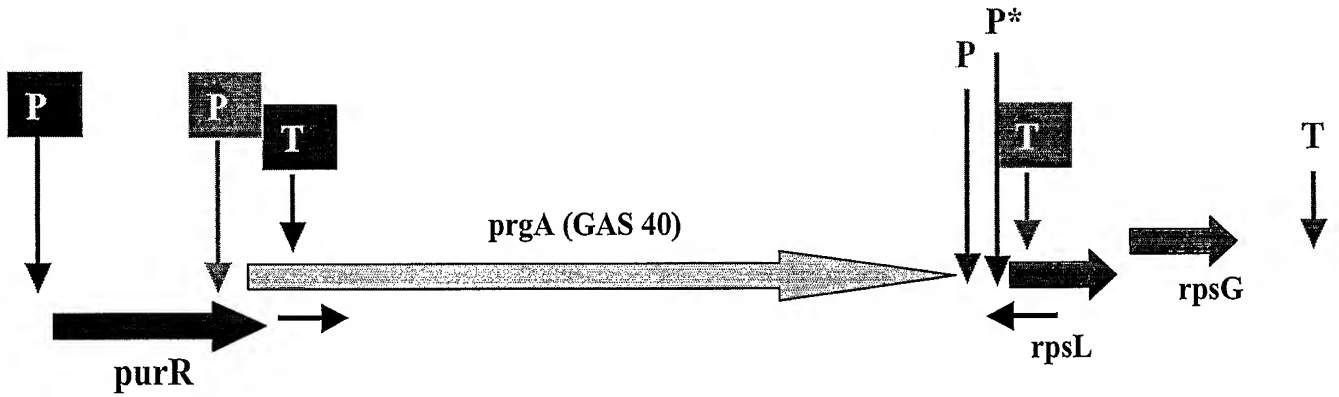
positions	58- 121	initial register 'f'	probability	0.811
positions	156- 196	initial register 'e'	probability	0.720
positions	198- 246	initial register 'e'	probability	1.000
positions	247- 261	initial register 'a'	probability	1.000
positions	556- 646	initial register 'e'	probability	1.000
positions	671- 733	initial register 'f'	probability	1.000

Figure 4(c): Secondary Structure prediction of Leucine Zipper within coiled coil.

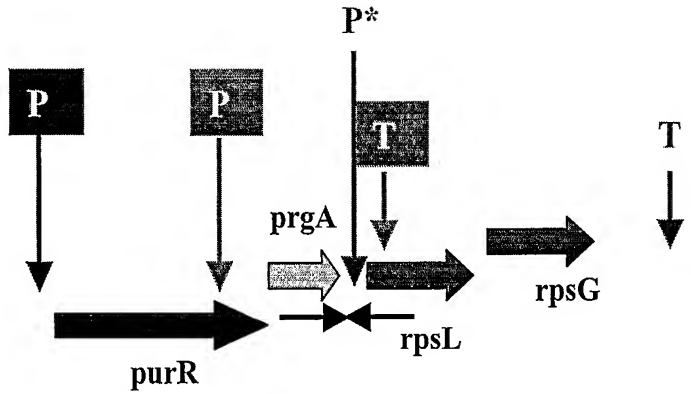
```
          673                                  701
QYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEH
            CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
            L------I------L------L------L
            OOOOOOOOOOOOOOOOOOOOOOOOOOOOO
```

Figure 5

5A
location of gas40 within GAS genome



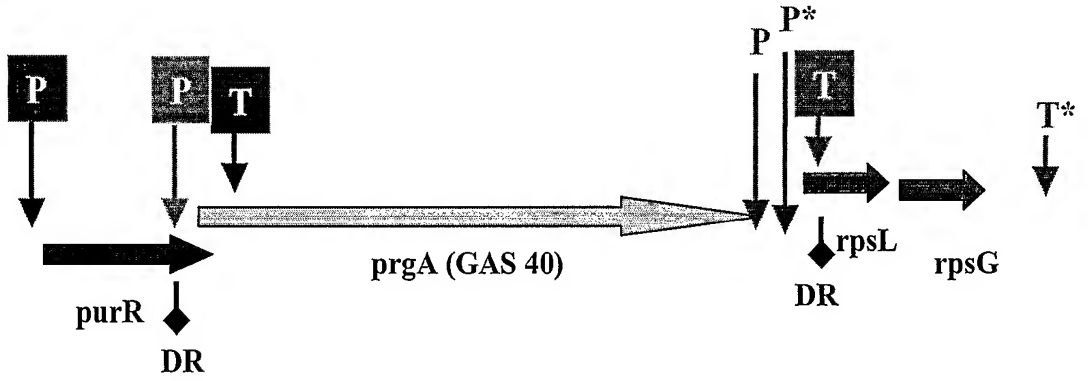
5B
Δgas40 deletion mutant



purine operon repressor	purR
GAS 40 surface exclusion protein	prgA
ribosomal protein S12	rpsL
ribosomal protein S7	rpsG
P	putative promoter
T	putative terminator

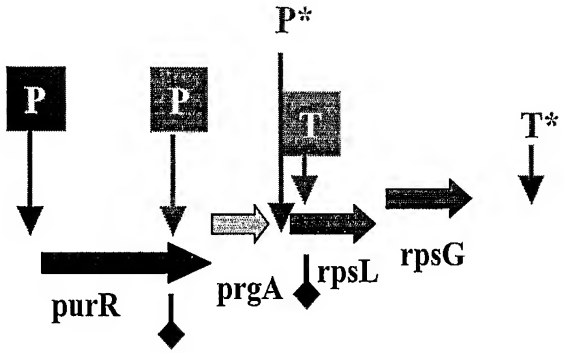
5C

direct repeats surrounding GAS 40 within wild type genome



5D

location of direct repeats within Δgas40



5E

corresponding genomic region within GBS

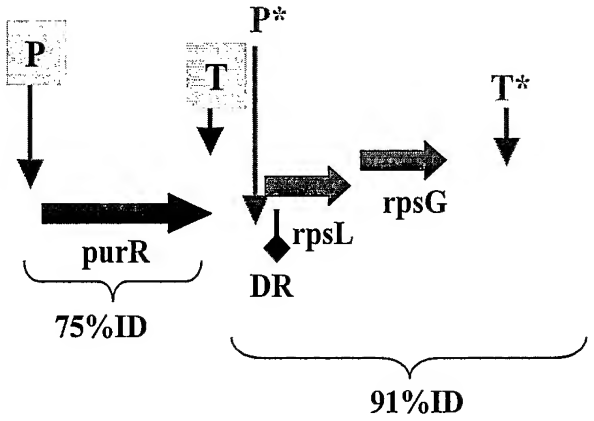
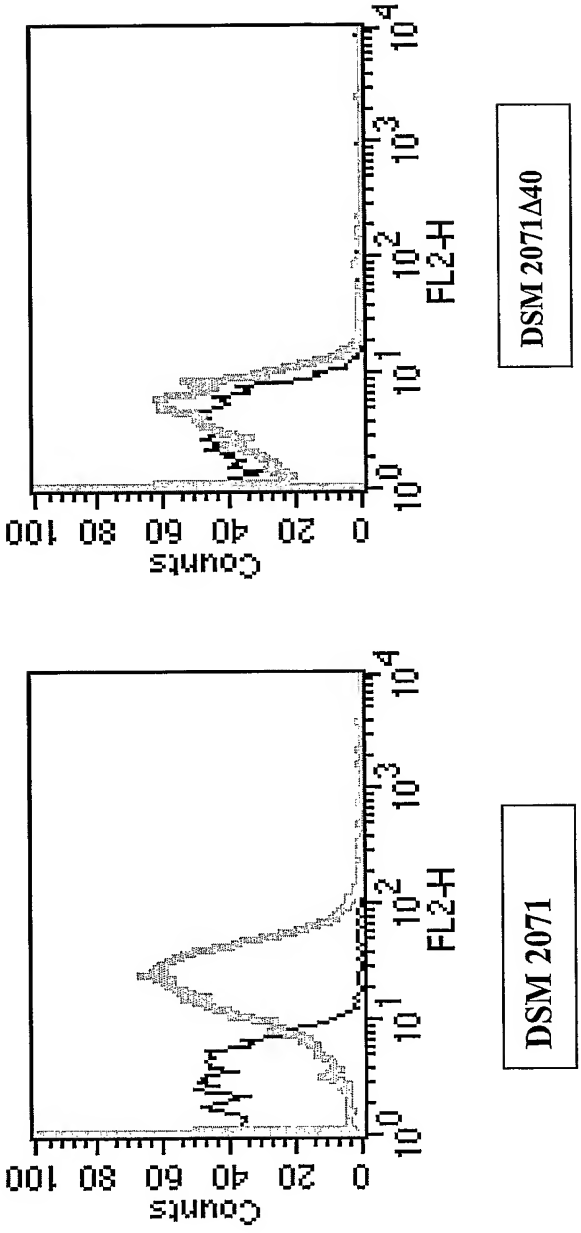
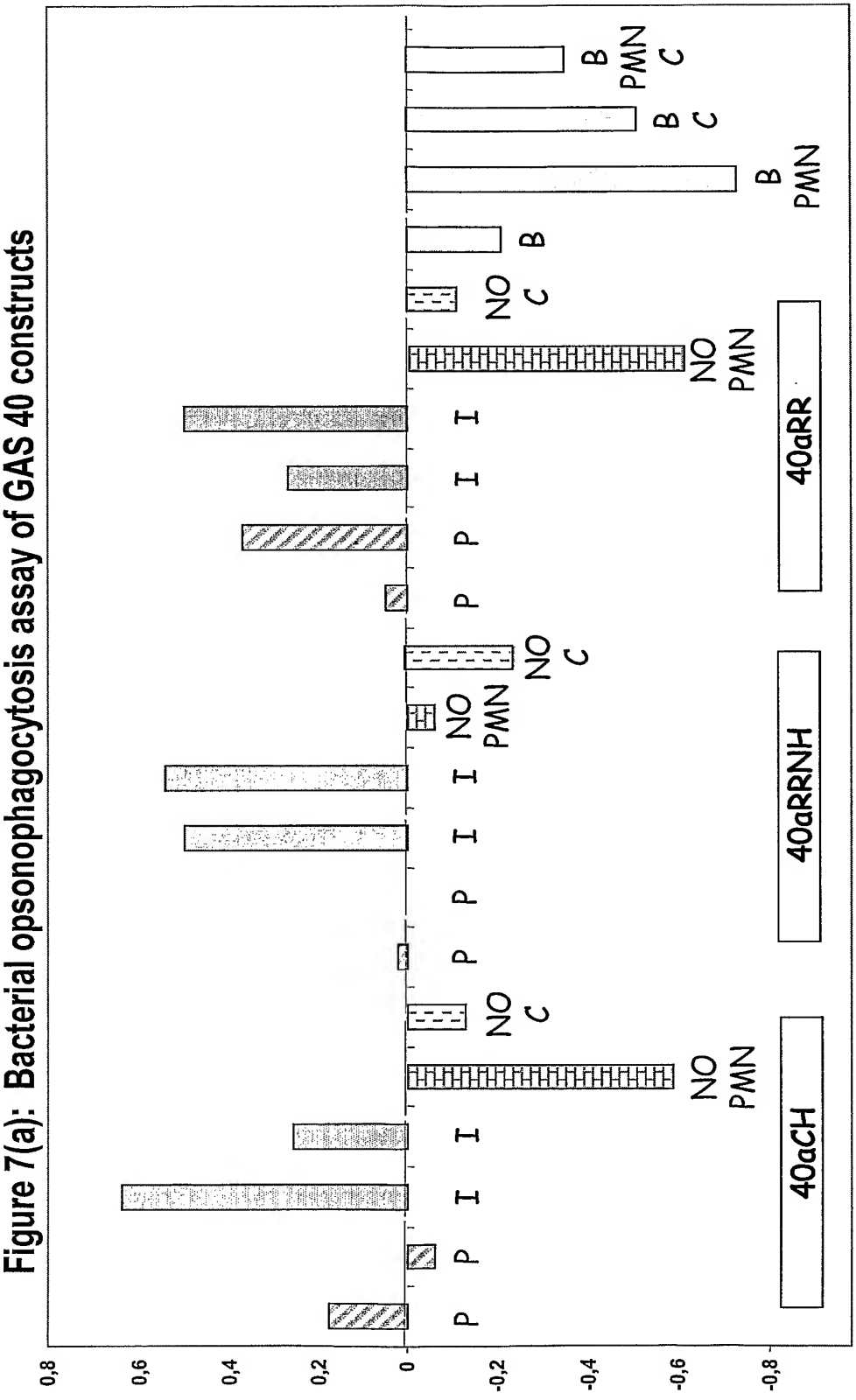


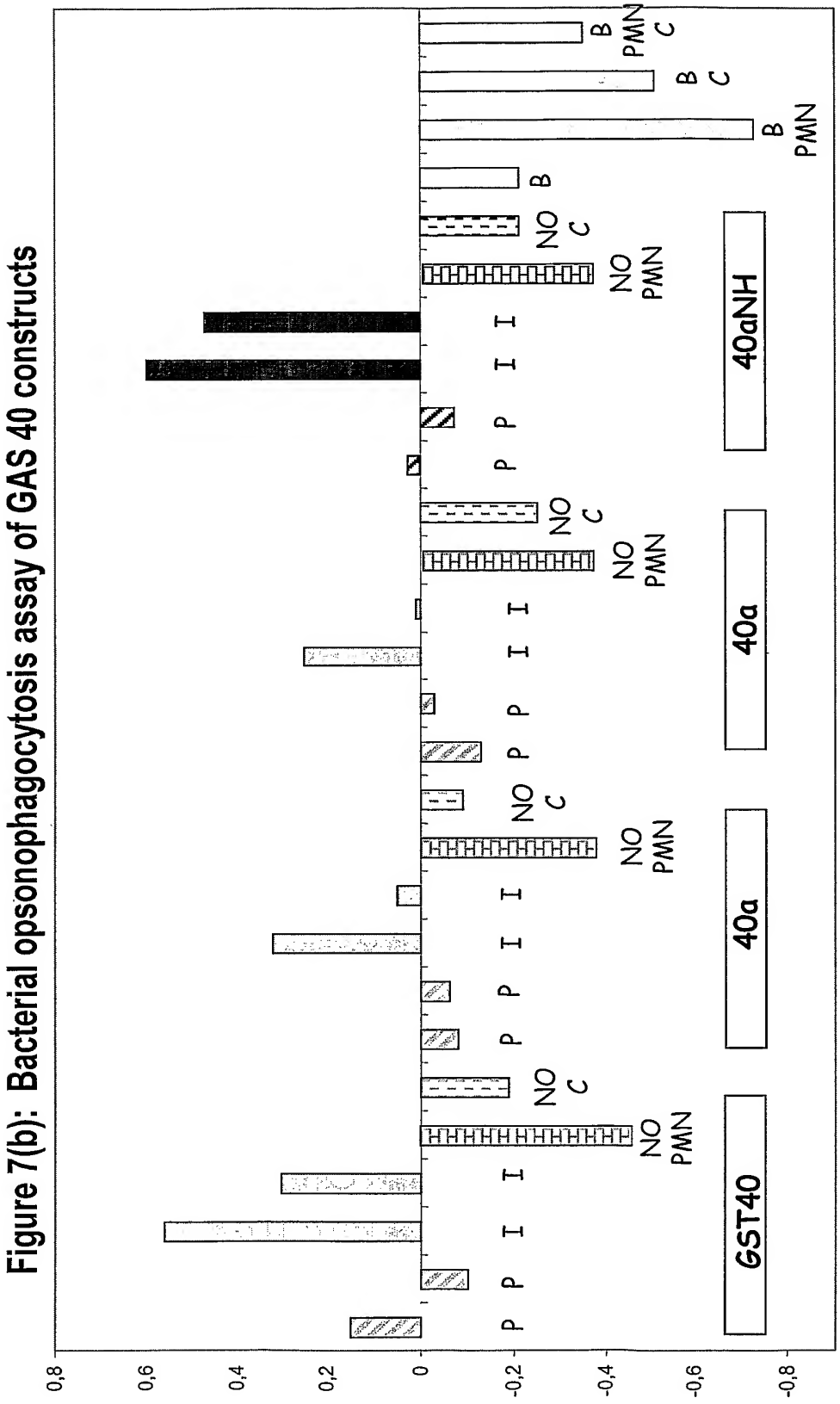
FIGURE 6: FACS Comparison of GAS 40 in wild type GAS and GAS 40 deletion mutant





B - bacteria
P - preimmune serum
I - immune serum
No PMN - without polymorphonucleates
No C - without complement

Histogram bars represent the difference between logarithm at T 0 (initial time) and T60 (bacterial CFU counted after 60 minutes of incubation)



Histogram bars represent the difference between logarithm at T 0 (initial time) and T60 (bacterial CFU counted after 60 minutes of incubation)

Figure 8: Immunization in Murine Mouse Model

GAS antigen	Survival/Tested mice			Protection	pValue	Protein Purity
	alive	dead	tested	%	Chi-square	%
gst 40	67	63	130	51	0.000012	
253	14	36	50	28	0.006	15
253-urea	2	8	10	20		25
253-gst	2	8	10	20		30
39	9	31	40	22.5	0.09	20
39a	13	37	50	26	0.016	10
39a	10	30	40	25	0.039	
39a	12	28	40	30	0.0046	
urea 366	21	78	99	21.2	0.046	65
117	19	51	70	27	0.0036	15
117-urea	1	9	10	10		80
117-urea-2M	7	23	30	23.3	0.1	80
117-urea-2M (prep 117)	8	32	40	20	0.2	
urea 504	9	31	40	22.5	0.09	50
504	14	26	40	35	0.0003	40
504	7	33	40	17.5	0.4	80
urea 389	7	23	30	23	0.1	30
533	14	56	70	20	0.12	50
new 533	4	16	20	20	0.34	30
gst 57	12	48	60	20	0.14	60
57a	0	20	20	0		50
294	17	73	90	18.8	0.14	80
130	15	65	80	18.7	0.17	40
130	7	23	30	23.3	0.1	40
84	8	32	40	20	0.2	70
urea 159	7	33	40	17.5	0.4	5
159a	2	8	10	20		65
527	10	40	50	20	0.17	50
527	3	17	20	15		80
217	7	33	40	17.5	0.4	50
511	13	67	80	16.2	0.41	80
277	8	42	50	16	0.52	5
277a	2	28	30	6.6		50
gst 202	3	17	20	10	0.75	5
202a	5	25	30	16.6	0.53	5
45	5	25	30	16.6	0.53	80
urea 309	5	25	30	20	0.53	8
290	6	34	40	15	0.67	50
529	6	34	40	15	0.67	5
gst 58	10	60	70	14.2	0.71	30
384	7	43	50	14	0.78	80
384RR	1	19	20	5		80
urea 509	7	53	60	11.6	0.84	50
509-NH	2	8	10			75
509-CH	0	10	10			75
193	7	53	60	11.6	0.84	65
urea 372	4	25	29	13.7	0.85	20
gst 42	4	26	30	13.3	0.9	50
95	5	35	40	12.5	1	55
urea 236	5	35	40	12.5	1	80
new 236	2	8	10	20		70
137	5	35	40	12.5	1	75
His-Stop	29	201	230	12:06		

SEQUENCE LISTING

SEQ ID NO: 1 amino acid sequence comprising GAS 40

MDLEQTKPNQVKQKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVE
KTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATE
TELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTKA
LSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSIVGNNTMKAPQGYPLEE
LKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPDNLTPEVQNELAQFAAHMI
NSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGPHDKTIIEDSAGASGLI
RNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLRVDKHNPNAPVYLGFSTSNV
GSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGKVSSLENRLSAIHQEADIMA
AQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLASLKAALHQT
EALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSS
LEATIATTEHQLTLLKTLANEKEYRHLDEDIATVPDLQVAPPLTGVKPLSYSKIDTTPLVQEMVKETKQLL
EASARLAAENTSLVAEALVGQTSEMVASNAIVSKITSSITQPSSKTSYGSGSSTTSNLISDVDESTQRALK
AGVVMLAAVGLTGFRFRKESK

SEQ ID NO: 2 polynucleotide sequence encoding for GAS 40

ATGGACTTAGAACAAACGAAGCCAAACCAAGTTAAGCAGAAAATTGCTTTAACCTCAACAATTGCTTTATT
GAGTGCCAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTA
ATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAA
AAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAAT
CAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATA
CTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAA
ACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCAT
TTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATATTGCTAAGC
TCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCA
TTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGAC
TGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAG
CTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAA
CTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGA
TCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATC
GCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATT
AATAGTGTAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATT
ACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCAGGGGTAT
CAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATT
CGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACG
TGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTA
TTAACTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTA
GGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGAC
CCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAG
CGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCTGATATTATGGCA
GCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCT
CCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAG
CACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACA
GAAGCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCT
AAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATT
TGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGT
CTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAATA
TCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAAAC
CGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACTATTA
GAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGGCCAAACCTCTGA
AATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTT
ATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAAGAGCTCTTAAA
GCAGGAGTCGTCATGTTGGCAGCTGTCGGCCTCACAGGATTTAGGTTCCGTAAGGAATCTAAGTGA

SEQ ID NO: 3 amino acid sequence comprising an N terminal leader sequence of GAS 40

MDLEQTKPNQVKQKIALTSTIALLSA

SEQUENCE LISTING

SEQ ID NO: 4 polynucleotide sequence encoding an N terminal leader sequence of GAS 40

ATGGACTTAGAACAAACGAAGCCAAACCAAGTTAAGCAGAAAATTGCTTTAACCTCAACAATTGCTTTATT
GAGTGCC

SEQ ID NO: 5 amino acid sequence comprising a fragment of GAS 40 with N terminal leader sequence removed

SVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINH
LKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATETELHNAQADQHSKETALSEQKASISA
ETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEE
LAAQKAALAEKEAELSRLKSSAPSTQDSIVGNNTMKAPQGYPLEELKKLEASGYIGSASYNNYYKEHADQI
IAKASPGNQLNQYQDIPADRNRFVDPDNLTPEVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLS
TSYKKTHGNTRPSFVYGQPGVSGHYGVGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGI
YDSIKYMLFTDHLHGNTYGHAINFLRVDKHNPNAPVYLGFSTSNVGSLNEHFVMFPESNIANHQRFNKTPI
KAVGSTKDYAQRVGTVSDTIAAIKGKVSSLENRLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQV
RQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRD
FKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRH
LDEDIATVPDLQVAPPLTGVKPLSYSKIDTTPLVQEMVKETKQLLEASARLAAENTSLVAEALVGQTSEMV
ASNAIVSKITSSITQPSSKTSYGSGSSTTSNLISDVDESTQRALKAGVVMLAAVGLTGFRFRKESK

SEQ ID NO: 6 polynucleotide sequence encoding a fragment of GAS 40 with N terminal leader sequence removed

AGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATACTCA
CGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAACTC
TCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAACCAC
TTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCTTGC
AAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAGAGC
TTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCATTTCAGCA
GAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATATTGCTAAGCTCAATGC
TATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAAGCT
CAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAAGAG
TTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAGCTCCGTC
TACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAACTTAAAA
AATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAAATT
ATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTTTGT
TGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAATAGTG
TAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATTACTTAGT
ACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCAGGGGTATCAGGGCA
TTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATTCGAAATG
ATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACGTGGTATT
TATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTATTAACTT
TTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTAGGATCTT
TGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCTATA
AAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAGCGATCAA
AGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCTGATATTATGGCAGCCCAAG
CTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAAGTG
AGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAGCACAACT
CGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAAGCCT
TAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCTAAGGGAC
TTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTTGGCTAA
AACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGTCTAGAAG
CTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAATATCGCCAC
TTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAAACCGCTATC
ATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACTATTAGAAGCTT
CAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGGCCAAACCTCTGAAATGGTA
GCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGGCTC

SEQUENCE LISTING

AGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAAGAGCTCTTAAAGCAGGAG
TCGTCATGTTGGCAGCTGTCGGCCTCACAGGATTTAGGTTCCGTAAGGAATCTAAGTGA

SEQ ID NO: 7 amino acid sequence comprising a C terminal transmembrane region of GAS 40
ALKAGVVMLAAVGLTGFRFRKESK

SEQ ID NO: 8 polynucleotide sequence encoding a C terminal transmembrane region of GAS 40
GCTCTTAAAGCAGGAGTCGTCATGTTGGCAGCTGTCGGCCTCACAGGATTTAGGTTCCGTAAGGAATCTAA
GTGA

SEQ ID NO: 9 amino acid sequence comprising a fragment of GAS 40 with a C terminal transmembrane sequence removed
MDLEQTKPNQVKQKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVE
KTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATE
TELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTKA
LSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSIVGNNTMKAPQGYPLEE
LKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPDNLTPEVQNELAQFAAHMI
NSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGPHDKTIIEDSAGASGLI
RNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLRVDKHNPNAPVYLGFSTSNV
GSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGKVSSLENRLSAIHQEADIMA
AQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLASLKAALHQT
EALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSS
LEATIATTEHQLTLLKTLANEKEYRHLDEDIATVPDLQVAPPLTGVKPLSYSKIDTTPLVQEMVKETKQLL
EASARLAAENTSLVAEALVGQTSEMVASNAIVSKITSSITQPSSKTSYGSGSSTTSNLISDVDESTQR

SEQ ID NO: 10 polynucleotide sequence encoding a fragment of GAS 40 with a C terminal transmembrane sequence removed
ATGGACTTAGAACAAACGAAGCCAAACCAAGTTAAGCAGAAAATTGCTTTAACCTCAACAATTGCTTTATT
GAGTGCCAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTA
ATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAA
AAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAAT
CAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATA
CTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAA
ACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCAT
TTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATATTGCTAAGC
TCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCA
TTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGAC
TGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAG
CTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAA
CTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGA
TCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATC
GCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATT
AATAGTGTAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATT
ACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCAGGGGTAT
CAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATT
CGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACG
TGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTA
TTAACTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTA
GGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGAC
CCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAG
CGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCTGATATTATGGCA
GCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCT
CCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAG
CACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACA
GAAGCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCT
AAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATT

SEQUENCE LISTING

TGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGT
CTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAATA
TCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAAAC
CGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACTATTA
GAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGGCCAAACCTCTGA
AATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTT
ATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAAGA

SEQ ID NO: 11 amino acid sequence comprising a transmembrane region of GAS 40 as shown in Figures 1 and 2. ALKAGVVMLAAVGLTG

SEQ ID NO: 12 amino acid sequence comprising a first coiled-coil region of GAS 40
ETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLL
AQGAEHQRELTATETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDA
ITKAAQTANDNTKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSA

SEQ ID NO: 13 amino acid sequence comprising a second coiled-coil region of GAS 40
RLSAIHQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSL
AKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNA
QEALAALQAKQSSLEATIATTEHQLTLLKTLANEKE

SEQ ID NO: 14 amino acid sequence comprising a leucine zipper motif within the second coiled-coil region of GAS 40.
QVIRERIDNTKQDLAKTTSSLLNAQEALAAL

SEQ ID NO: 15 amino acid sequence comprising SpA from *Streptococcus gordonii* Genbank reference GI 25990270
MNKRKEVFGFRKSKVAKTLCGAVLGAALIAIADQQVLADEVTETNSTANVAVTTTGNPATNLPEAQGEATE
AASQSQAQAGSKEGALPVEVSADDLNQAVTDAKAAGVNVVQDQTSDKGTATTAAENAQKQAEIKSDYAKQA
EEIKKTTEAYKKEVEAHQAETDKINAENKAAEDKYQEDLKAHQAEVEKINTANATAKAEYEAKLAQYQKDL
AAVQKANEDSQLDYQNKLSAYQAELARVQKANAEAKEAYEKAVKENTAKNAALQAENEAIKQRNETAKANY
DAAMKQYEADLAAIKKAKEDNDADYQAKLAAYQAELARVQKANADAKAAYEKAVEENTAKNTAIQAENEAI
KQRNAAAKATYEAALKQYEADLAAAKKANEDSDADYQAKLAAYQTELARVQKANADAKAAYEKAVEDNKAK
NAALQAENEEIKQRNAAAKTDYEAKLAKYEADLAKYKKELAEYPAKLKAYEDEQAQIKAALVELEKNKNQD
GYLSKPSAQSLVYDSEPNAQLSLTTNGKMLKASAVDEAFSHDTAQYSKKILQPDNLNVSYLQQADDVTSSM
ELYGNFGDKAGWTTTVGNNTEVKFASVLLERGQSVTATYTNLEKSYYNGKKISKAVFKYSLDSDSKFKNVD
KAWLGVLPDPTLGVFASAYTGQEEKDTSIFIKNEFTFYDENDQPINFDNALLSVASLNRENNSIEMAKDYS
GTFVKISGSSVGEKDGKIYATETLNFKQGQGGSRWTMYKNSQPGSGWDSSDAPNSWYGAGAISMSGPTNHV
TVGAISATQVVPSDPVMAVATGKRPNIWYSLNGKIRAVNVPKITKEKPTPPVAPTEPQAPTYEVEKPLEPA
PVAPTYENEPTPPVKTPDQPEPSKPEEPTYETEKPLEPAPVVPTYENEPTPPVKTPDQPEPSKPEEPTYET
EKPLEPAPVAPTYENEPTPPVKTPDQPEPSKPEEPTYDPLPTPPVAPTPKQLPTPPVVPTVHFHYSSLLAQ
PQINKEIKNEDGVDIDRTLVAKQSIVKFELKTEALTAGRPKTTSFVLVDPLPTGYKFDLDATKAASTGFDT
TYDEASHTVTFKATDETLATYNADLTKPVETLHPTVVGRVLNDGATYINNFTLTVNDAYGIKSNVVRVTTP
GKPNDPDNPNNNYIKPTKVNKNKEGLNIDGKEVLAGSTNYYELTWDLDQYKGDKSSKEAIQNGFYYVDDYP
EEALDVRPDLVKVADEKGNQVSGVSVQQYDSLEAAPKKVQDLLKKANITVKGAFQLFSADNPEEFYKQYVS
TGTSLVITDPMTVKSEFGKTGGKYENKAYQIDFGNGYATEVVVNNVPKITPKKDVTVSLDPTSENLDGQTV
QLYQTFNYRLIGGFIPQNHSEELEDYSFVDDYDQAGDQYTGNYKTFSSLNLTMKDGSVIKAGTDLTSQTTA
ETDAANGIVTVRSKEDSLQKISLDSPFQAETYLQMRRIAIGTFENTYVNTVNKVAYASNTVRTTTPIPRTP
DKPTPIPTPKPKDPDKPETPKEPKVPSPKVEDPSAPIPVSVGKELTTLPKTGTNDSSYMPYLGLAALVGVL
GLGQLKRKEDESN

SEQ ID NO: 16 amino acid sequence comprising Streptococcal surface protein B precursor from *Streptococcus gordonii* Genbank reference GI 25055226 AAC44102.3

SEQUENCE LISTING

MQKREVFGFRKSKVAKTLCGAVLGAALIAIADQQVLADEVTETNSTANVAVTTTGNPATNLPEAQGEATEA
ASQSQAQAGSKDGALPVEVSADDLNKAVTDAKAAGVNVVQDQTSDKGTATTAAENAQKQAEIKSDYAKQAE
EIKKTTEAYKKEVEAHQAETDKINAENKAAEDKYQEDLKAHQAEVEKINTANATAKAEYEAKLAQYQKDLA
AVQKANEDSQLDYQNKLSAYQAELARVQKANAEAKEAYEKAVKENTAKNAALQAENEAIKQRNETAKANYD
AAMKQYEADLAAIKKAKEDNDADYQAKLAAYQAELARVQKANADAKAAYEKAVEENTAKNTAIQAENEAIK
QRNETAKATYEAAVKQYEADLAAVKQANATNEADYQAKLAAYQTELARVQKANADAKATYEKAVEDNKAKN
AALQAENEEIKQRNAAAKTDYEAKLAKYEADLAKYKKDFAAYTAALAEAESKKKQDGYLSEPRSQSLNFKS
EPNAIRTIDSSVHQYGQQELDALVKSWGISPTNPDRKKSTAYSYFNAINSNNTYAKLVLEKDKPVDVTYTG
LKNSSFNGKKISKVVYTYTLKETGFDDGTKMTMFASSDPTVTAWYNDYFTSTNINVKVKFYDEEGQLMNLT
GGLVNFSSLNRGNGSGAIDKDAIESVRNFNGRYIPISGSSIKIHENNSAYADSSNAEKSRGARWDTSEWDT
TSSPNNWYGAIVGEITQSEISFNMASSKSGNIWFAFNSNINAIGVPTKPVAPTAPTQPMYETEKPLEPAPV
VPTYENEPTPPVKTPDQPEPSKPEEPTYETEKPLEPAPVAPTYENEPTPPVKIPDQPEPSKPEEPTYETEK
PLEPAPVAPTYENEPTPPVKTPDQPEPSKPEEPTYDPLPTPPLAPTPKQLPTPPVVPTVHFHYSSLLAQPQ
INKEIKNEDGVDIDRTLVAKQSIGKFELKTEALTAGRPKTTSFVLVDPLPTGYKFDLDATKAASTGFDTTY
DEASHTVTFKATDETLATYNADLTKPVETLHPTVVGRVLNDGATYTNNFTLTVNDAYGIKSNVVRVTTPGK
PNDPDNPNNNYIKPTKVNKNKEGLNIDGKEVLAGSTNYYELTWDLDQYKGDKSSKEAIQNGFYYVDDYPEE
ALDVRPDLVKVADEKGNQVSGVSVQQYDSLEAAPKKVQDLLKKANITVKGAFQLFSADNPEEFYKQYVSTG
TSLVITDPMTVKSEFGKTGGKYENKAYQIDFGNGYATEVVVNNVPKITPKKDVTVSLDPTSENLDGQTVQL
YQTFNYRLIGGFIPQNHSEELEDYSFVDDYDQAGDQYTGNYKTFSSLNLTMKDGSVIKAGTDLTSQTTAET
DATNGIVTVRFKEDFLQKISLDSPFQAETYLQMRRIAIGTFENTYVNTVNKVAYASNTVRTTTPIPRTPDK
PTPIPTPKPKDPDKPETPKEPKVPSPKVEDPSAPIPVSVGKELTTLPKTGTNDATYMPYLGLAALVGFLGL
GLAKRKED

SEQ ID NO: 17 amino acid sequence comprising PspA from *Streptococcus pneumoniae* Genbank reference GI 282335

MNKKKMILTSLASVAILGAGFVASQPTVVRAEESPVASQSKAEKDYDAAKKDAKNAKKAVEDAQKALDDAK
AAQKKYDEDQKKTEEKAALEKAASEEMDKAVAAVQQAYLAYQQATDKAAKDAADKMIDEAKKREEEAKTKF
NTVRAMVVPEPEQLAETKKKSEEAKQKAPELTKKLEEAKAKLEEAEKKATEAKQKVDAEEVAPQAKIAELE
NQVHRLEQELKEIDESESEDYAKEGFRAPLQSKLDAKKAKLSKLEELSDKIDELDAEIAKLEDQLKAAEEN
NNVEDYFKEGLEKTIAAKKAELEKTEADLKKAVNEPEKPAPAPETPAPEAPAEQPKPAPAPQPAPAPKPEK
PAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYFYNTDGSMATGWLQ
NNGSWYYLNSNGAMATGWLQYNGSWYYLNANGAMATGWAKVNGSWYYLNANGAMATGWLQYNGSWYYLNAN
GAMATGWAKVNGSWYYLNANGAMATGWLQYNGSWYYLNANGAMATGWAKVNGSWYYLNANGAMATGWVKDG
DTWYYLEASGAMKASQWFKVSDKWYYVNGLGALAVNTTVDGYKVNANGEWV

SEQ ID NO: 18 amino acid sequence comprising a portion of Se89.9 of *Streptococcus equi* Genbank reference GI 2330384

ESDIVDATRFSTTEIPKSGQVIDRSASIQALTNDIASIKGKIASLESRLADPSSEAEVTAAQAKISQLQH
QLEAAQAKSHKLDQQVEQLANTKDSLRTQLLAAKEEQAQLKANLDKALALLASSKATLHKLEAAMEEAKA
RVAGLASQKAQLEDLLAFEKNPNRIELAQEKVAAAKKALADTEDKLLAAQASLSDLQAQRARLQLSIATI

SEQ ID NO: 19 polynucleotide sequence comprising GST-40-HIS

CTGGTTCCGCGTGGATCCCATATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGG
AGAAACGAAGGCGAGTAATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAA
CTATTGATGCAGTTGAAAAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACA
AAAACTACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACA
AGAAATTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAG
AGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCA
GAACAAAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGA
ACAAAATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTA
ATGATAATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAA
GTTAAAAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAG
TCGTCTTAAATCCTCAGCTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAG
GCTATCCTCTTGAAGAACTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTAT
TACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATAT
TCCAGCAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGT

SEQUENCE LISTING

TTGCAGCTCACATGATTAATAGTGTAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCA
CAAGAATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTA
CGGACAGCCAGGGGTATCAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCG
GAGCGTCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACT
GTGAATGGTATTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAA
TACATACGGCCATGCTATTAACTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGAT
TTTCAACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCAT
CAACGCTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGT
ATCTGATACTATTGCAGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAG
AAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAG
TCAGACAGCTTAAATCTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGC
AGCTAAAGCAAAACAAGCACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAG
CCGCACTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAA
GCTCATTTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGA
TAATACTAAGCAAGATTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTAC
AAGCTAAACAAAGCAGTCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTA
GCTAACGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACC
TCTTACGGGCGTAAAACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAG
AAACGAAACAACTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTT
GTTGGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCC
CTCATCTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTA
CTCAAAGAGCTCTTAAAGCAGGAGTCGTCATGTTGGCAGCTGTCGGCCTCACAGGATTTAGGTTCCGTAAG
GAATCTAAG**GCGGCCGCACTCGAGCACCACCACCACCACCACCAC**

SEQ ID NO: 20 amino acid sequence comprising GST-40-HIS

L V P R G S H Met S V G V S H Q V K A D D R A S G E T K A S N T H D D S
L P K P E T I Q E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T
K T T A E I N H L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E
T L L A Q G A E H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E
Q K A S I S A E T T R A Q D L V E Q V K T S E Q N I A K L N A Met I S N
P D A I T K A A Q T A N D N T K A L S S E L E K A K A D L E N Q K A K V
K K Q L T E E L A A Q K A A L A E K E A E L S R L K S S A P S T Q D S I
V G N N T Met K A P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y
K E H A D Q I I A K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L
T P E V Q N E L A Q F A A H Met I N S V R R Q L G L P P V T V T A G S Q
E F A R L L S T S Y K K T H G N T R P S F V Y G Q P G V S G H Y G V G P
H D K T I I E D S A G A S G L I R N D D N Met Y E N I G A F N D V H T V
N G I K R G I Y D S I K Y Met L F T D H L H G N T Y G H A I N F L R V D
K H N P N A P V Y L G F S T S N V G S L N E H F V Met F P E S N I A N H
Q R F N K T P I K A V G S T K D Y A Q R V G T V S D T I A A I K G K V S
S L E N R L S A I H Q E A D I Met A A Q A K V S Q L Q G K L A S T L K Q
S D S L N L Q V R Q L N D T K G S L R T E L L A A K A K Q A Q L E A T R
D Q S L A K L A S L K A A L H Q T E A L A E Q A A A R V T A L V A K K A
H L Q Y L R D F K L N P N R L Q V I R E R I D N T K Q D L A K T T S S L
L N A Q E A L A A L Q A K Q S S L E A T I A T T E H Q L T L L K T L A N
E K E Y R H L D E D I A T V P D L Q V A P P L T G V K P L S Y S K I D T
T P L V Q E Met V K E T K Q L L E A S A R L A A E N T S L V A E A L V G
Q T S E Met V A S N A I V S K I T S S I T Q P S S K T S Y G S G S S T T
S N L I S D V D E S T Q R A L K A G V V Met L A A V G L T G F R F R K E
S K **A A A L E H H H H H H**

SEQ ID NO: 21 polynucleotide sequence comprising 40a-HIS

ATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATAC
TCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAA
CTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAAC
CACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCT

SEQUENCE LISTING

TGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAG
AGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCATTTCA
GCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATATTGCTAAGCTCAA
TGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAA
GCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAA
GAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAGCTCC
GTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAACTTA
AAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAA
ATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTT
TGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAATA
GTGTAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATTACTT
AGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCAGGGGTATCAGG
GCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATTCGAA
ATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACGTGGT
ATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTATTAA
CTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTAGGAT
CTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCT
ATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAGCGAT
CAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCTGATATTATGGCAGCCC
AAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAA
GTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAGCACA
ACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAAG
CCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCTAAGG
GACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTTGGC
TAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGTCTAG
AAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAATATCGC
CACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAAACCGCT
ATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACTATTAGAAG
CTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGGCCAAACCTCTGAAATG
GTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGG
CTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAA**cGtGCGGCCGCACTCG**
AGCACCACCACCACCACCACCAC

SEQ ID NO: 22 amino acid sequence comprising 40a-HIS

M S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T I Q
E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H
L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E
H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E
T T R A Q D L V E Q V K T S E Q N I A K L N A Met I S N P D A I T K A A
Q T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E L
A A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T M K A
P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A
K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A
Q F A A H M I N S V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y
K K T H G N T R P S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A
G A S G L I R N D D N M Y E N I G A F N D V H T V N G I K R G I Y D S I
K Y M L F T D H L H G N T Y G H A I N F L R V D K H N P N A P V Y L G F
S T S N V G S L N E H F V M F P E S N I A N H Q R F N K T P I K A V G S
T K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E A
D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T
K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L
H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N R
L Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q
S S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T V
P D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E M V K E T K Q L
L E A S A R L A A E N T S L V A E A L V G Q T S E M V A S N A I V S K I

SEQUENCE LISTING

T S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q R **A A A L E H H H H H H**

SEQ ID NO: 23 polynucleotide sequence comprising 40a-RR-HIS

ATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATAC
TCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAA
CTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAAC
CACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCT
TGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAG
AGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCATTTCA
GCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATATTGCTAAGCTCAA
TGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAA
GCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAA
GAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAGCTCC
GTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAACTTA
AAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAA
ATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTT
TGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAATA
GTGTA**cGtcGt**CAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATTACTT
AGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCAGGGGTATCAGG
GCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATTCGAA
ATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACGTGGT
ATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTATTAA
CTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTAGGAT
CTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCT
ATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAGCGAT
CAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCTGATATTATGGCAGCCC
AAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAA
GTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAGCACA
ACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAAG
CCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCTAAGG
GACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTTGGC
TAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGTCTAG
AAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAATATCGC
CACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAAACCGCT
ATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACTATTAGAAG
CTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGGCCAAACCTCTGAAATG
GTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGG
CTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAA**cGtGCGGCCGCACTCG**
AGCACCACCACCACCACCACCAC

SEQ ID NO: 24 amino acid sequence comprising 40a-RR-HIS

M S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T I Q
E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H
L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E
H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E
T T R A Q D L V E Q V K T S E Q N I A K L N A Met I S N P D A I T K A A
Q T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E L
A A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T M K A
P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A
K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A
Q F A A H M I N S V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y
K K T H G N T R P S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A
G A S G L I R N D D N M Y E N I G A F N D V H T V N G I K R G I Y D S I
K Y M L F T D H L H G N T Y G H A I N F L R V D K H N P N A P V Y L G F
S T S N V G S L N E H F V M F P E S N I A N H Q R F N K T P I K A V G S

SEQUENCE LISTING

T K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E A
D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T
K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L
H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N R
L Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q
S S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T V
P D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E M V K E T K Q L
L E A S A R L A A E N T S L V A E A L V G Q T S E M V A S N A I V S K I
T S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q R **A A A**
L E H H H H H H H

SEQ ID NO: 25 polynucleotide sequence comprising 40a-RR (nat)

ATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATAC
TCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAA
CTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAAC
CACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCT
TGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAG
AGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCATTTCA
GCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATATTGCTAAGCTCAA
TGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAA
GCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAA
GAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAGCTCC
GTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAACTTA
AAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAA
ATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTT
TGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAATA
GTGTA**cGtcGt**CAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATTACTT
AGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCAGGGGTATCAGG
GCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATTCGAA
ATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACGTGGT
ATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTATTAA
CTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTAGGAT
CTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCT
ATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAGCGAT
CAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCTGATATTATGGCAGCCC
AAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAA
GTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAGCACA
ACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAAG
CCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCTAAGG
GACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTTGGC
TAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGTCTAG
AAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAATATCGC
CACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAAACCGCT
ATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACTATTAGAAG
CTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGGCCAAACCTCTGAAATG
GTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGG
CTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAA**cGt**

SEQ ID NO: 26 amino acid sequence comprising 40a-RR (nat)

M S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T I Q
E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H
L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E
H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E
T T R A Q D L V E Q V K T S E Q N I A K L N A M I S N P D A I T K A A Q
T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E L A
A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T M K A P

SEQUENCE LISTING

Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K
A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A Q
F A A H M I N S V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y K
K T H G N T R P S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A G
A S G L I R N D D N M Y E N I G A F N D V H T V N G I K R G I Y D S I K
Y M L F T D H L H G N T Y G H A I N F L R V D K H N P N A P V Y L G F S
T S N V G S L N E H F V M F P E S N I A N H Q R F N K T P I K A V G S T
K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E A D
I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T K
G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L H
Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N R L
Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q S
S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T V P
D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E M V K E T K Q L L
E A S A R L A A E N T S L V A E A L V G Q T S E M V A S N A I V S K I T
S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q R

SEQ ID NO: 27 polynucleotide sequence comprising HIS-40a NH

ATGGGATCGCATCACCATCACCATCACGCTAGTAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAG
AGCCTCAGGAGAAACGAAGGCGAGTAATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGG
CAAAGGCAACTATTGATGCAGTTGAAAAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACC
GCTCTGACAAAAACTACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAAC
CTCTGCACAAGAAATTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAA**C**
ACCAAAGAGAGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACT
GCATTGTCAGAACAAAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAA
AACGTCTGAACAAAATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTC
AAACGGCTAATGATAATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAA
AAAGCTAAAGTTAAAAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGC
AGAACTTAGTCGTCTTAAATCCTCAGCTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAG
CACCGCAAGGCTATCCTCTTGAAGAACTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTAC
AATAATTATTACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATA
CCAAGATATTCCAGCAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGC
TAGCGCAGTTTGCAGCTCACATGATTAATAGTGTAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACA
GCAGGATCACAAGAATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATC
ATTTGTCTACGGACAGCCAGGGGTATCAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAG
ACTCTGCCGGAGCGTCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGAT
GTGCATACTGTGAATGGTATTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTT
ACACGGAAATACATACGGCCATGCTATTAACTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTT
ACCTTGGATTTTCAACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATT
GCTAACCATCAACGCTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAA**GAC**TATGCCCAAAGAGT
AGGCACTGTATCTGATACTATTGCAGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTA
TTCATCAAGAAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACA
CTTAAGCAGTCAGACAGCTTAAATCTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGA
ATTACTAGCAGCTAAAGCAAAACAAGCACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCAT
CGTTGAAAGCCGCACTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTG
GCTAAAAAAGCTCATTTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGA
GCGCATTGATAATACTAAGCAAGATTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAG
CAGCCTTACAAGCTAAACAAAGCAGTCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTT
AAAACCTTAGCTAACGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGT
AGCTCCACCTCTTACGGGCGTAAAACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAA
TGGTTAAAGAAACGAAACAACTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCA
GAAGCGCTTGTTGGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGAT
TACTCAGCCCTCATCTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTG
ATGAAAGTACTCAA**cGt**

SEQ ID NO: 28 amino acid sequence comprising HIS-40a NH

SEQUENCE LISTING

M G S H H H H H H A S S V G V S H Q V K A D D R A S G E T K A S N T H D
D S L P K P E T I Q E A K A T I D A V E K T L S Q Q K A E L T E L A T A
L T K T T A E I N H L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S
E E T L L A Q G A E H Q R E L T A T E T E L H N A Q A D Q H S K E T A L
S E Q K A S I S A E T T R A Q D L V E Q V K T S E Q N I A K L N A M I S
N P D A I T K A A Q T A N D N T K A L S S E L E K A K A D L E N Q K A K
V K K Q L T E E L A A Q K A A L A E K E A E L S R L K S S A P S T Q D S
I V G N N T M K A P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y
K E H A D Q I I A K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L
T P E V Q N E L A Q F A A H M I N S V R R Q L G L P P V T V T A G S Q E
F A R L L S T S Y K K T H G N T R P S F V Y G Q P G V S G H Y G V G P H
D K T I I E D S A G A S G L I R N D D N M Y E N I G A F N D V H T V N G
I K R G I Y D S I K Y M L F T D H L H G N T Y G H A I N F L R V D K H N
P N A P V Y L G F S T S N V G S L N E H F V M F P E S N I A N H Q R F N
K T P I K A V G S T K D Y A Q R V G T V S D T I A A I K G K V S S L E N
R L S A I H Q E A D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N
L Q V R Q L N D T K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A
K L A S L K A A L H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L
R D F K L N P N R L Q V I R E R I D N T K Q D L A K T T S S L L N A Q E
A L A A L Q A K Q S S L E A T I A T T E H Q L T L L K T L A N E K E Y R
H L D E D I A T V P D L Q V A P P L T G V K P L S Y S K I D T T P L V Q
E M V K E T K Q L L E A S A R L A A E N T S L V A E A L V G Q T S E M V
A S N A I V S K I T S S I T Q P S S K T S Y G S G S S T T S N L I S D V
D E S T Q R

SEQ ID NO: 29 polynucleotide sequence comprising HIS-40a CH

ATGGCTAGTAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAG
TAATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTG
AAAAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTG**ACG**AAAACTACTGCTGAA
ATCAAC**CAT**TTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAA
TACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTG
AAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGC
ATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATATTGCTAA
GCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAG
CATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTG
ACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTC
AGCTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAG
AACTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCA
GATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAA
TCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGA
TTAATAGTGTAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGA
TTACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCA**TCT**GTCTACGGACAGCCAGGGGT
ATCAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCA
TTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAA
CGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGC
TATTAACTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATG
TAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAG
ACCCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGC
AGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCTGATATTATGG
CAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAAT
CTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACA
AGCACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGA
CAGAAGCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATAT
CTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGA
TTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCA
GTCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAA

SEQUENCE LISTING

TATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAA
ACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACTAT
TAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGGCCAAACCTCT
GAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATC
TTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAA**cGtGCGGCCG**
CACTCGAGCACCACCACCACCACCAC

SEQ ID NO: 30 amino acid sequence comprising HIS-40a CH

M A S S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T
I Q E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I
N H L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G
A E H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S
A E T T R A Q D L V E Q V K T S E Q N I A K L N A M I S N P D A I T K A
A Q T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E
L A A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T M K
A P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I
A K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L
A Q F A A H M I N S V R R Q L G L P P V T V T A G S Q E F A R L L S T S
Y K K T H G N T R P S S V Y G Q P G V S G H Y G V G P H D K T I I E D S
A G A S G L I R N D D N M Y E N I G A F N D V H T V N G I K R G I Y D S
I K Y M L F T D H L H G N T Y G H A I N F L R V D K H N P N A P V Y L G
F S T S N V G S L N E H F V M F P E S N I A N H Q R F N K T P I K A V G
S T K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E
A D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D
T K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A
L H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N
R L Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K
Q S S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T
V P D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E M V K E T K Q
L L E A S A R L A A E N T S L V A E A L V G Q T S E M V A S N A I V S K
I T S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q **R A A**
A L E H H H H H H

SEQ ID NO: 31 polynucleotide sequence comprising HIS-40a-RR NH

ATGGGATCGCATCACCATCACCATCACGCTAGTAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAG
AGCCTCAGGAGAAACGAAGGCGAGTAATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGG
CAAAGGCAACTATTGATGCAGTTGAAAAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACC
GCTCTGACAAAAACTACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAAC
CTCTGCACAAGAAATTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAAC
ATCAAAGAGAGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACT
GCATTGTCAGAACAAAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAA
AACGTCTGAACAAAATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTC
AAACGGCTAATGATAATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAA
AAAGCTAAAGTTAAAAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGC
AGAACTTAGTCGTCTTAAATCCTCAGCTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAG
CACCGCAAGGCTATCCTCTTGAAGAACTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTAC
AATAATTATTACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATA
CCAAGATATTCCAGCAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGC
TAGCGCAGTTTGCAGCTCACATGATTAATAGTGTA**cGtcGt**CAATTAGGTCTACCACCAGTTACTGTTACA
GCAGGATCACAAGAATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATC
ATTTGTCTACGGACAGCCAGGGGTATCAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAG
ACTCTGCCGGAGCGTCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGAT
GTGCATACTGTGAATGGTATTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTT
ACACGGAAATACATACGGCCATGCTATTAACTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTT
ACCTTGGATTTTCAACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATT
GCTAACCATCAACGCTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGT

SEQUENCE LISTING

AGGCACTGTATCTGATACTATTGCAGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTA
TTCATCAAGAAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACA
CTTAAGCAGTCAGACAGCTTAAATCTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGA
ATTACTAGCAGCTAAAGCAAAACAAGCACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCAT
CGTTGAAAGCCGCACTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTG
GCTAAAAAAGCTCATTTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGA
GCGCATTGATAATACTAAGCAAGATTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAG
CAGCCTTACAAGCTAAACAAAGCAGTCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTT
AAAACCTTAGCTAACGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGT
AGCTCCACCTCTTACGGGCGTAAAACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAA
TGGTTAAAGAAACGAAACAACTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCA
GAAGCGCTTGTTGGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGAT
TACTCAGCCCTCATCTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTG
ATGAAAGTACTCAA**cGt**

SEQ ID NO: 32 amino acid sequence comprising HIS-40a-RR NH

M G S H H H H H H A S S V G V S H Q V K A D D R A S G E T K A S N T H D
D S L P K P E T I Q E A K A T I D A V E K T L S Q Q K A E L T E L A T A
L T K T T A E I N H L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S
E E T L L A Q G A E H Q R E L T A T E T E L H N A Q A D Q H S K E T A L
S E Q K A S I S A E T T R A Q D L V E Q V K T S E Q N I A K L N A M I S
N P D A I T K A A Q T A N D N T K A L S S E L E K A K A D L E N Q K A K
V K K Q L T E E L A A Q K A A L A E K E A E L S R L K S S A P S T Q D S
I V G N N T M K A P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y
K E H A D Q I I A K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L
T P E V Q N E L A Q F A A H M I N S V R R Q L G L P P V T V T A G S Q E
F A R L L S T S Y K K T H G N T R P S F V Y G Q P G V S G H Y G V G P H
D K T I I E D S A G A S G L I R N D D N M Y E N I G A F N D V H T V N G
I K R G I Y D S I K Y M L F T D H L H G N T Y G H A I N F L R V D K H N
P N A P V Y L G F S T S N V G S L N E H F V M F P E S N I A N H Q R F N
K T P I K A V G S T K D Y A Q R V G T V S D T I A A I K G K V S S L E N
R L S A I H Q E A D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N
L Q V R Q L N D T K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A
K L A S L K A A L H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L
R D F K L N P N R L Q V I R E R I D N T K Q D L A K T T S S L L N A Q E
A L A A L Q A K Q S S L E A T I A T T E H Q L T L L K T L A N E K E Y R
H L D E D I A T V P D L Q V A P P L T G V K P L S Y S K I D T T P L V Q
E M V K E T K Q L L E A S A R L A A E N T S L V A E A L V G Q T S E M V
A S N A I V S K I T S S I T Q P S S K T S Y G S G S S T T S N L I S D V
D E S T Q R

SEQ ID NO: 33 polynucleotide sequence comprising 40N-HIS

ATGCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATACTCACGACGATAGTTTACC
AAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAACTCTCAGTCAACAAAAAG
CAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAACCACTTAAAAGAGCAGCAA
GATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCTTGCAAGTAGTGAGGAGAC
GCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAG
CAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCATTTCAGCAGAAACTACTCGAGCT
CAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATATTGCTAAGCTCAATGCTATGATTAGCAATCC
TGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAAGCTCAGAATTGGAGAAGG
CTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAA
GCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAGCTCCGTCTACTCAAGATAGCAT
TGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAACTTAAAAAATTAGAAGCTAGTG
GTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAGT
CCAGGTAATCAATTAAATCAATACCAA**GCGGCCGCACTCGAGCACCACCACCACCACCACCAC**

SEQUENCE LISTING

SEQ ID NO: 34 amino acid sequence comprising 40N-HIS

MQVKADDRASGETKASNTHDDSLPKPETIQEAKATI
DAVEKTLSQQKAELTELATALTKTTAEINHLKEQQD
NEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELT
ATETELHNAQADQHSKETALSEQKASISAETTRAQD
LVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNT
KALSSELEKAKADLENQKAKVKKQLTEELAAQKAAL
AEKEAELSRLKSSAPSTQDSIVGNNTMKAPQGYPLE
ELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQ
LNQYQAAALEHHHHHHHH

SEQ ID NO: 35 amino acid sequence comprising GAS 117

MTLKKHYYLLSLLALVTVGAAFNTSQSVSAQVYSNEGYHQHLTDEKSHLQYSKDNAQLQLRNILDGYQNDL
GRHYSSYYYYNLRTVMGLSSEQDIEKHYEELKNKLHDMYNHY

SEQ ID NO: 36 polynucleotide sequence encoding GAS 117

ATGACACTAAAAAAACACTATTATCTTCTCAGCCTGCTAGCTCTTGTAACGGTTGGTGCTGCCTTTAACAC
AAGCCAGAGTGTCAGTGCACAAGTTTATAGCAATGAAGGGTATCACCAGCATTTGACTGATGAAAAATCAC
ACCTGCAATATAGTAAAGACAACGCACAACTTCAATTGAGAAATATCCTTGACGGCTACCAAAATGACCTA
GGGAGACACTACTCTAGCTATTATTACTACAACCTAAGAACCGTTATGGGACTATCAAGTGAGCAAGACAT
TGAAAAACACTATGAAGAGCTTAAGAACAAGTTACATGATATGTACAATCATTATTAA

SEQ ID NO: 37 amino acid sequence comprising GAS 117 leader sequence

TLKKHYYLLSLLALVTVGA

SEQ ID NO: 38 amino acid sequence comprising fragment of GAS 117 where leader sequence is removed

AFNTSQSVSAQVYSNEGYHQHLTDEKSHLQYSKDNAQLQLRNILDGYQNDLGRHYSSYYYYNLRTVMGLSS
EQDIEKHYEELKNKLHDMYNHY

SEQ ID NO: 39 amino acid sequence comprising GAS 130

MSHMKKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGIDYAHARGAKVYVAAN
MVTHEGNEIGAGEWFRQLRDMGLDAVIVSDPALIVICSTEAPGLEIHLSTQASSTNYETFEFWKAMGLTRV
VLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLSNHMSHRDANRGGCSQSCRWKYDLYDMPFGG
ERRSLKGEIPEDYSMSSVDMCMIDHIPDLIENGVDSLKIEGRMKSIHYVSTVTNCYKAAVGAYMESPEAFY
AIKEELIDELWKVAQRELATGFYYGIPTENEQLFGARRKIPQYKFVGEVVAFDSASMTATIRQRNVIMEGD
RIECYGPGFRHFETVVKDLHDADGQKIDRAPNPMELLTISLPREVKPGDMIRACKEGLVNLYQKDGTSKTV
RT

SEQ ID NO: 40 polynucleotide sequence encoding GAS 130

ATGTCACATATGAAAAAACGTCCCGAGGTCTTATCACCTGCTGGAACACTTGAAAAATTAAAAGTTGCGAT
TGACTATGGCGCAGATGCTGTTTTTGTTGGAGGGCAGGCCTATGGCCTAAGAAGCCGCGCTGGTAACTTCT
CTATGGAAGAATTGCAAGAAGGCATTGATTATGCACATGCGCGTGGAGCTAAGGTCTATGTTGCTGCTAAC
ATGGTTACCCACGAAGGGAACGAAATTGGTGCGGGCGAGTGGTTTCGTCAACTGCGTGATATGGGGCTTGA
TGCGGTCATTGTTTCAGATCCAGCCTTGATTGTTATTTGTTCAACAGAAGCCCCAGGTTTGGAAATTCATT
TGTCAACGCAAGCTTCATCTACCAATTACGAGACCTTTGAATTTTGGAAAGCCATGGGCTTGACCCGAGTT
GTTTTAGCTCGCGAGGTTAATATGGCCGAGTTAGCAGAAATCCGCAAGCGGACAGATGTGGAAATTGAAGC
CTTTGTCCATGGAGCCATGTGTATCTCTTATTCAGGCCGCTGTGTTTTGTCAAACCACATGAGTCACCGTG
ATGCCAACAGGGGCGGCTGCTCACAGTCTTGCCGCTGGAAGTATGATTTGTATGACATGCCATTTGGAGGA
GAGCGCCGCTCCTTAAAAGGGGAAATTCCAGAAGACTATTCTATGTCCTCTGTTGACATGTGTATGATTGA
CCATATTCCTGACCTGATTGAAAATGGGGTTGATAGCTTAAAAATTGAAGGCCGAATGAAATCTATCCACT
ACGTCTCAACCGTAACCAACTGTTACAAGGCGGCTGTAGGTGCTTACATGGAAAGCCCAGAAGCTTTTTAT
GCTATCAAAGAGGAATTGATTGACGAGTTGTGGAAGGTTGCCCAGCGCGAGTTGGCTACAGGTTTTTACTA
TGGTATCCCAACTGAAAATGAACAATTATTTGGTGCTCGCCGCAAAATTCCACAATATAAATTTGTCGGAG
AAGTAGTTGCCTTTGACTCAGCTAGCATGACAGCGACCATTCGTCAGCGTAATGTCATCATGGAAGGCGAT

SEQUENCE LISTING

CGGATTGAATGTTATGGACCAGGTTTCCGTCATTTTGAAACGGTTGTTAAGGACTTACATGATGCGGATGG
CCAAAAGATTGACCGTGCCCCAAATCCAATGGAACTCTTAACCATCTCTTTACCGAGAGAAGTTAAGCCAG
GGGATATGATTAGGGCTTGCAAGGAAGGTCTGGTTAACCTCTATCAAAAAGATGGCACCAGTAAAACTGTT
AGAACATAG

SEQ ID NO: 41 amino acid sequence comprising GAS 277

MTTMQKTISLLSLALLIGLLGTSGKAISVYAQDQHTDNVIAESTISQVSVEASMRGTEPYIDATVTTDQPV
RQPTQATITLKDASDNTINSWVYTMAAQQRRFTAWFDLTGQKSGDYHVTVTVHTQEKAVTGQSGTVHFDQN
KARKTPTNMQQKDTSKAMTNSVDVDTKAQTNQSANQEIDSTSNPFRSATNHRSTSLKRSTKNEKLTPTASN
SQKNGSNKTKMLVDKEEVKPTSKRGFPWVLLGLVVSLAAGLFIAIQKVSRRK

SEQ ID NO: 42 polynucleotide sequence encoding GAS 277

ATGACAACTATGCAAAAAACAATTAGCTTATTATCACTAGCTTTACTTATTGGTTTGCTGGGGACTTCTGG
CAAAGCCATATCTGTGTATGCACAAGATCAGCACACTGATAATGTTATAGCTGAATCAACTATTAGTCAGG
TCAGTGTTGAAGCCAGTATGCGTGGAACAGAACCTTATATTGATGCTACAGTCACCACAGATCAACCTGTC
AGACAACCAACTCAGGCAACGATAACACTTAAAGACGCTAGTGATAATACTATTAATAGTTGGGTATATAC
TATGGCAGCGCAACAGCGTCGTTTTACAGCTTGGTTTGATTTAACTGGACAAAAGAGTGGTGACTATCATG
TAACTGTCACCGTTCATACTCAAGAAAAGGCAGTAACTGGTCAATCAGGAACTGTTCATTTTGATCAAAAC
AAAGCTAGAAAAACACCAACTAATATGCAACAAAAGGATACTTCTAAAGCAATGACGAATTCAGTCGATGT
AGACACAAAAGCTCAAACAAATCAATCAGCTAACCAAGAAATAGATTCTACTTCAAATCCTTTCAGATCAG
CTACTAATCATCGATCAACTTCCTTAAAGCGATCTACTAAAAATGAGAAACTTACACCAACTGCTAGTAAT
AGCCAAAAAAACGGTAGCAACAAGACAAAAATGCTAGTGGACAAAGAGGAAGTAAAACCTACTTCAAAAAG
AGGATTCCCTTGGGTCTTATTAGGTCTAGTAGTCAGTTTAGCTGCAGGTTTATTTATAGCTATTCAAAAAG
TATCTAGACGAAAATAA

SEQ ID NO: 43 amino acid sequence comprising N-terminal leader sequence of GAS 277

TTMQKTISLLSLALLIGLLGTSGKAISVYA

SEQ ID NO: 44 amino acid sequence comprising fragment of GAS 277 where N-terminal leader sequence is removed

QDQHTDNVIAESTISQVSVEASMRGTEPYIDATVTTDQPVRQPTQATITLKDASDNTINSWVYTMAAQQRR
FTAWFDLTGQKSGDYHVTVTVHTQEKAVTGQSGTVHFDQNKARKTPTNMQQKDTSKAMTNSVDVDTKAQTN
QSANQEIDSTSNPFRSATNHRSTSLKRSTKNEKLTPTASNSQKNGSNKTKMLVDKEEVKPTSKRGFPWVLL
GLVVSLAAGLFIAIQKVSRRK

SEQ ID NO: 45 amino acid sequence comprising GAS 236

MTQMNYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGMLLSAFHMYEKELD
KVRFVGIHTGHLGFYTDYRDFEVDKLIDNLRKDKGEQISYPILKVAITLDDGRVVKARALNEATVKRIEKT
MVADVIINHVKFESFRGDGISVSTPTGSTAYNKSLGGAVLHPTIEALQLTEISSLNNRVFRTLGSSIIIPK
KDKIELVPKRLGIYTISIDNKTYQLKNVTKVEYFIDDEKIHFVSSPSHTSFWERVKDAFIGEIDS

SEQ ID NO: 46 polynucleotide sequence encoding GAS 236

ATGACACAGATGAATTATACAGGTAAGGTAAAACGAGTTGCTATTATTGCAAATGGTAAGTACCAAAGTAA
ACGCGTCGCCTCCAAACTTTTCTCCGTATTTAAAGATGATCCTGATTTCTATCTTTCAAAGAAAAATCCGG
ATATTGTGATTTCTATTGGCGGAGATGGGATGCTCTTATCTGCCTTTCACATGTATGAAAAAGAATTAGAT
AAGGTACGTTTTGTAGGAATCCACACCGGTCATCTTGGCTTTTATACCGATTATAGGGATTTTGAAGTTGA
TAAATTAATTGATAATTTAAGAAAAGACAAGGGAGAACAAATCTCTTATCCGATTTTAAAAGTTGCTATTA
CTTTAGATGATGGTCGTGTGGTTAAAGCGCGTGCTTTGAATGAAGCGACGGTTAAGCGTATTGAAAAAACG
ATGGTAGCAGATGTTATTATTAACCATGTCAAATTTGAAAGCTTCCGAGGTGATGGGATTTCAGTATCGAC
CCCGACAGGGAGCACAGCCTACAATAAATCTTTAGGTGGTGCTGTCTTGCATCCGACGATTGAAGCGCTGC
AATTGACGGAAATTTCCAGTCTTAATAACCGTGTCTTTAGAACCTTGGGCTCATCAATCATTATTCCCAAA
AAAGATAAGATTGAGTTAGTGCCAAAACGATTAGGAATTTATACCATTTCCATTGATAATAAAACCTATCA
GTTAAAAAATGTGACGAAGGTGGAGTATTTTATCGACGATGAGAAAATTCATTTTGTTTCCTCTCCGAGTC
ATACGAGCTTTTGGGAAAGGGTCAAGGATGCCTTTATTGGAGAGATTGACTCATGA

SEQUENCE LISTING

SEQ ID NO: 47 amino acid sequence comprising N-terminus leader sequence of GAS 236

MTQM

SEQ ID NO: 48 amino acid sequence comprising a fragment of GAS 236 where the N-terminal leader sequence is removed

NYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGMLLSAFHMYEKELDKVRF
VGIHTGHLGFYTDYRDFEVDKLIDNLRKDKGEQISYPILKVAITLDDGRVVKARALNEATVKRIEKTMVAD
VIINHVKFESFRGDGISVSTPTGSTAYNKSLGGAVLHPTIEALQLTEISSLNNRVFRTLGSSIIIPKKDKI
ELVPKRLGIYTISIDNKTYQLKNVTKVEYFIDDEKIHFVSSPSHTSFWERVKDAFIGEIDS

SEQ ID NO: 49 amino acid sequence comprising GAS 389

MRNEMAKIMNVTGEEVIALAATYMTKADVAFVAKALAYATAAHFYQVRKSGEPYIVHPIQVAGILADLHLD
AVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQLAENHRKMLMAMSKDIRVI
LVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGISRIKWELEDLAFRYLNETEFYKISHMMKE
KRREREALVEAIVSKVKTYTTQQGLFGDVYGRPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYA
MVGYIHELWRPMPGRFKDYIAAPKANGYQSIHTTVYGPKGPIEIQIRTKDMHQVAEYGVAAHWAYKKGVRG
KVNQAEQAVGMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDFAYAIHT
QIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIRQFFKNQDKELSVNK
GRDLLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFGDISPISVFNKLTEKERREEERAKA
KAEAEELVKGGEVKHENKDVLKVRSENGVIIQGASGLLMRIAKCCNPVPGDPIDGYITKGRGIAIHRSDCH
NIKSQDGYQERLIEVEWDLDNSSKDYQAEIDIYGLNRSGLLNDVLQILSNSTKSISTVNAQPTKDMKFANI
HVSFGIPNLTHLTTVVEKIKAVPDVYSVKRTNG

SEQ ID NO: 50 polynucleotide sequence encoding GAS 389

ATGAGGAACGAAATGGCAAAAATAATGAACGTAACAGGAGAAGAAGTCATTGCCTTAGCGGCCACCTATAT
GACCAAGGCTGATGTGGCTTTTGTGGCAAAGGCTTTAGCATATGCAACAGCGGCCCATTTCTACCAAGTGA
GAAAGTCAGGCGAACCCTATATCGTCCATCCGATTCAGGTGGCGGGGATTCTGGCTGATTTGCATCTGGAT
GCTGTGACAGTTGCTTGTGGCTTTTTACATGATGTCGTAGAAGATACGGATATTACCTTAGATGAGATCGA
AGCAGACTTTGGCCATGATGCTCGTGATATCGTTGATGGTGTCACCAAGTTAGGTGAAGTTGAGTACAAAT
CTCATGAGGAGCAACTCGCCGAAAACCATCGCAAAATGCTGATGGCTATGTCCAAAGATATTCGCGTGATT
TTGGTGAAATTGGCTGACCGCCTGCATAATATGCGCACCCTCAAACATTTGCGCAAGGACAAACAAGAGCG
CATTTCGCGCGAAACCATGGAAATCTATGCCCCCTTGGCGCATCGTTTGGGGATTAGTCGCATCAAATGGG
AACTAGAAGATTTGGCTTTTCGTTACCTCAATGAAACCGAATTTTACAAAATTTCCCATATGATGAAAGAA
AAACGTCGCGAGCGTGAAGCTTTGGTAGAGGCTATTGTCAGTAAGGTCAAAACCTATACGACACAACAAGG
GTTGTTTGGAGATGTGTATGGCCGACCAAAACACATTTATTCGATTTATCGGAAAATGCGGGACAAAAAGA
AACGATTCGATCAGATTTTTGATCTGATTGCCATTCGTTGTGTCATGGAAACGCAAAGCGATGTCTATGCT
ATGGTTGGCTATATTCATGAGCTTTGGCGTCCCATGCCAGGCCGCTTCAAGGATTATATTGCAGCTCCTAA
AGCTAATGGCTACCAGTCTATTCATACCACCGTGTATGGGCCAAAAGGACCTATTGAGATTCAAATCAGAA
CTAAGGACATGCATCAAGTGGCTGAGTACGGGGTTGCTGCTCACTGGGCTTATAAAAAAGGCGTGCGTGGT
AAGGTCAATCAAGCTGAGCAAGCCGTTGGCATGAACTGGATCAAAGAGCTGGTAGAATTGCAAGATGCCTC
AAATGGCGATGCAGTGGACTTTGTGGATTCGGTCAAAGAAGACATTTTTTCTGAACGGATTTATGTCTTTA
CACCGACAGGGGCCGTTCAGGAGTTACCAAAAGAATCAGGTCCTATTGATTTTGCTTATGCGATCCATACG
CAAATCGGTGAAAAAGCAACAGGTGCCAAAGTCAATGGACGTATGGTTCCTCTCACTGCCAAGTTAAAAAC
AGGAGATGTGGTTGAAATCATCACCAATGCCAATTCCTTTGGCCCTAGTCGAGACTGGGTAAAACTGGTCA
AAACCAATAAGGCTCGCAACAAAATTCGTCAGTTCTTTAAAAATCAAGACAAGGAATTGTCAGTGAATAAA
GGCCGTGATTTGTTGGTGTCTTATTTTCAAGAGCAGGGCTACGTTGCCAATAAATACCTTGACAAAAAACG
CATTGAAGCCATCCTTCCAAAAGTCAGTGTGAAGAGCGAAGAATCACTCTATGCAGCCGTTGGGTTTGGTG
ACATTAGTCCTATCAGTGTCTTTAACAAGTTAACCGAAAAAGAGCGCCGTGAAGAAGAAAGGGCCAAGGCT
AAAGCAGAAGCTGAAGAATTGGTTAAGGGCGGTGAGGTCAAACACGAAAACAAAGATGTGCTCAAGGTTCG
CAGTGAAAATGGAGTCATTATCCAAGGAGCATCAGGCCTCTTGATGCGGATTGCCAAGTGTTGTAATCCTG
TACCTGGTGATCCTATTGACGGCTACATTACCAAAGGGCGTGGCATTGCGATTCACAGATCGGACTGTCAT
AACATTAAGAGTCAAGATGGCTACCAAGAACGCTTGATTGAGGTCGAGTGGGATTTGGACAATTCGAGTAA
AGATTATCAGGCTGAAATTGATATCTATGGGCTCAATCGTAGTGGTCTGCTTAATGATGTGCTCCAAATTT
TATCAAACTCAACCAAGAGCATATCGACAGTCAATGCTCAGCCGACCAAGGACATGAAGTTTGCTAATATT
CACGTGAGCTTTGGCATTCCAAATCTGACGCATCTGACCACTGTTGTCGAAAAAATCAAGGCAGTTCCAGA
TGTTTATAGCGTGAAGCGGACCAATGGCTAA

SEQUENCE LISTING

SEQ ID NO: 51 amino acid sequence comprising GAS 504

MKTRITELLNIDYPIFQGGMAWVADGDLAGAVSNAGGLGIIGGGNAPKEVVKANIDRVKAITDRPFGVNIM
LLSPFADDIVDLVIEEGVKVVTTGAGNPGKYMERLHQAGIIVVPVVPSVALAKRMEKLGVDAVIAEGMEAG
GHIGKLTTMSLVRQVVEAVSIPVIAAGGIADGHGAAAAFMLGAEAVQIGTRFVVAKESNAHQNFKDKILAA
KDIDTVISAQVVGHPVRSIKNKLTSAYAKAEKAFLIGQKTATDIEEMGAGSLRHAVIEGDVVNGSVMAGQI
AGLVRKEESCETILKDIYYGAARVIQNEAKRWQSVSIEK

SEQ ID NO: 52 polynucleotide sequence encoding GAS 504

ATGAAAACACGTATTACAGAATTACTTAATATTGATTACCCCATTTTTCAAGGAGGAATGGCTTGGGTTGC
TGATGGTGATTTAGCAGGTGCAGTTTCTAATGCTGGTGGTTTAGGCATTATAGGTGGTGGCAATGCTCCCA
AAGAAGTCGTTAAAGCTAATATTGATCGTGTCAAAGCTATTACTGATAGACCTTTTGGGGTTAATATCATG
CTTTTATCTCCTTTTGCTGATGATATCGTTGATCTGGTCATTGAAGAAGGTGTTAAAGTAGTAACAACAGG
CGCAGGAAATCCAGGAAAGTATATGGAAAGACTGCACCAGGCGGGTATAATCGTTGTTCCTGTTGTCCCAA
GCGTTGCGCTAGCCAAACGTATGGAAAAGCTTGGGGTAGATGCTGTTATTGCTGAGGGTATGGAAGCTGGA
GGACATATTGGCAAGTTAACGACTATGTCTTTAGTAAGACAAGTTGTTGAAGCGGTTTCGATTCCTGTCAT
TGCGGCAGGTGGTATAGCTGATGGTCATGGTGCAGCAGCAGCATTTATGTTAGGAGCAGAGGCTGTTCAAA
TTGGAACTCGCTTTGTTGTTGCTAAAGAATCCAATGCTCACCAAAATTTTAAAGATAAAATCTTAGCAGCA
AAAGATATTGATACGGTGATTTCTGCGCAGGTTGTGGGCCACCCTGTCCGTTCTATTAAAAATAAATTGAC
CTCAGCTTACGCTAAAGCAGAAAAAGCATTTTTAATTGGTCAAAAAACAGCTACTGATATTGAAGAAATGG
GAGCAGGATCGCTTCGACACGCTGTTATTGAAGGCGATGTAGTCAATGGATCTGTTATGGCTGGCCAAATT
GCAGGGCTTGTGAGAAAAGAAGAAAGCTGTGAAACGATTTTAAAAGATATTTATTATGGTGCAGCTCGTGT
TATTCAAAATGAAGCTAAGCGCTGGCAATCTGTTTCAATAGAAAAGTAG

SEQ ID NO: 53 amino acid sequence comprising GAS 509

MTKIYKTITELVGQTPIIKLNRLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEGLISPGDVIIE
PTSGNTGIGLAWVGAAKGYRVIIVMPETMSLERRQIIQAYGAELVLTPGAEGMKGAIAKAETLAIELGAW
MPMQFNNPANPSIHEKTTAQEILEAFKEISLDAFVSGVGTGGTLSGVSHVLKKANPETVIYAVEAEESAV
LSGQEPGPHKIQGISAGFIPNTLDTKAYDQIIRVKSKDALETARLTGAKEGFLVGISSGAALYAAIEVAK
QLGKGKHVLTILPDNGERYLSTELYDVPVIKTK

SEQ ID NO: 54 polynucleotide sequence encoding GAS 509

ATGACTAAAATTTACAAAACTATAACAGAATTAGTAGGTCAAACACCTATTATCAAACTTAACCGTTTAA
TTCCAAACGAAGCTGCTGACGTTTATGTAAAATTAGAAGCTTTTAACCCAGGATCTTCTGTTAAAGATCG
TATTGCTTTATCGATGATTGAAGCTGCTGAAGCTGAAGGTCTGATAAGTCCTGGTGACGTTATTATCGAA
CCAACAAGTGGTAATACAGGTATTGGTCTTGCATGGGTAGGTGCTGCTAAAGGGTATCGAGTCATTATTG
TTATGCCCGAAACTATGAGCTTGGAAAGACGGCAAATCATTCAGGCTTATGGTGCAGAGCTTGTCTTAAC
ACCTGGAGCAGAAGGTATGAAAGGGGCTATTGCAAAAGCTGAAACTTTAGCAATAGAACTAGGTGCTTGG
ATGCCTATGCAATTTAATAACCCTGCCAATCCAAGCATCCATGAAAAAACAACAGCTCAAGAAATTTTGG
AAGCTTTTAAGGAGATTTCTTTAGATGCATTCGTATCTGGTGTTGGTACTGGAGGAACACTTTCTGGTGT
TTCACATGTCTTGAAAAAAGCTAACCCTGAAACTGTTATCTATGCTGTTGAAGCTGAAGAATCTGCTGTC
TTATCTGGTCAAGAGCCTGGACCACATAAAATTCAAGGTATATCAGCTGGATTTATCCCAAACACGTTAG
ATACCAAAGCCTATGACCAAATTATCCGTGTTAAATCGAAAGATGCTTTAGAAACTGCTCGACTAACAGG
AGCTAAGGAAGGCTTCCTGGTTGGGATTTCTTCTGGAGCTGCTCTTTACGCCGCTATTGAAGTCGCTAAA
CAGTTAGGAAAAGGCAAACATGTGTTAACTATTTTACCAGATAATGGCGAACGCTATTTATCGACTGAAC
TCTATGATGTACCAGTAATTAAGACGAAATAA

SEQ ID NO: 55 amino acid sequence comprising C-terminus transmembrane region of GAS 509

FLVGISSGAALYAAIEVAKQLGKGKHVLTILPDNGERYLSTELYDVPVIKTK

SEQ ID NO: 56 amino acid sequencing comprising a fragment of GAS 509 where the C-terminal transmembrane region is removed

MTKIYKTITELVGQTPIIKLNRLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEGLISPGDVIIEP
TSGNTGIGLAWVGAAKGYRVIIVMPETMSLERRQIIQAYGAELVLTPGAEGMKGAIAKAETLAIELGAWMP

SEQUENCE LISTING

MQFNNPANPSIHEKTTAQEILEAFKEISLDAFVSGVGTGGTLSGVSHVLKKANPETVIYAVEAEESAVLSG
QEPGPHKIQGISAGFIPNTLDTKAYDQIIRVKSKDALETARLTGAKEG

SEQ ID NO: 57 amino acid sequence comprising GAS 366

MKVISNFQNKKILILGLAKSGEAAAKLLTKLGALVTVNDSKPFDQNPAAQALLEEGIKVICGSHPVELLDE
NFEYMVKNPGIPYDNPMVKRALAKEIPILTEVELAYFVSEAPIIGITGSNGKTTTTTMIADVLNAGGQSAL
LSGNIGYPASKVVQKAIAGDTLVMELSSFQLVGVNAFRPHIAVITNLMPTHLDYHGSFEDYVAAKWMIQAQ
MTESDYLILNANQEISATLAKTTKATVIPFSTQKVVDGAYLKDGILYFKEQAIIAATDLGVPGSHNIENAL
ATIAVAKLSGIADDIIAQCLSHFGGVKHRLQRVGQIKDITFYNDSKSTNILATQKALSGFDNSRLILIAGG
LDRGNEFDDLVPDLLGLKQMIILGESAERMKRAANKAEVSYLEARNVAEATELAFKLAQTGDTILLSPANA
SWDMYPNFEVRGDEFLATFDCLRGDA

SEQ ID NO: 58 polynucleotide sequence encoding GAS 366

ATGAAAGTGATAAGTAATTTTCAAAACAAAAAAATATTAATATTGGGGTTAGCCAAATCGGGCGAAGCAGC
AGCAAAATTATTGACCAAACTTGGTGCTTTAGTGACTGTTAATGATAGTAAACCATTTGACCAAAATCCAG
CGGCACAAGCCTTGTTGGAAGAGGGGATTAAGGTCATTTGTGGTAGCCACCCAGTAGAATTATTAGATGAG
AACTTTGAGTACATGGTTAAAAACCCTGGGATTCCTTATGATAATCCTATGGTTAAACGCGCCCTTGCAAA
GGAAATTCCCATCTTGACTGAAGTAGAATTGGCTTATTTCGTATCTGAAGCGCCTATTATCGGGATTACAG
GATCAAACGGGAAGACAACCACAACGACAATGATTGCCGATGTTTTGAATGCTGGCGGGCAATCTGCACTC
TTATCTGGAAACATTGGTTATCCTGCTTCAAAAGTTGTTCAAAAAGCAATTGCTGGTGATACTTTGGTGAT
GGAATTGTCCTCTTTTCAATTAGTGGGAGTGAATGCTTTTCGCCCTCATATTGCTGTCATCACTAATTTAA
TGCCGACTCACCTGGACTATCATGGCAGTTTTGAGGATTATGTTGCTGCTAAATGGATGATTCAAGCTCAG
ATGACAGAATCAGACTACCTTATTTTAAATGCTAATCAAGAGATTTCAGCAACTCTAGCTAAGACCACCAA
AGCAACAGTGATTCCTTTTTCAACTCAAAAAGTGGTTGATGGAGCTTATCTGAAGGATGGAATACTCTATT
TTAAAGAACAGGCGATTATAGCTGCAACTGACTTAGGTGTCCCAGGTAGCCACAACATTGAAAATGCCCTA
GCAACTATTGCAGTTGCCAAGTTATCTGGTATTGCTGATGATATTATTGCCCAGTGCCTTTCACATTTTGG
AGGCGTTAAACATCGTTTGCAACGGGTTGGTCAAATCAAAGATATTACCTTCTACAATGACAGTAAGTCAA
CCAATATTTTAGCCACTCAAAAAGCTTTATCAGGTTTTGATAACAGTCGCTTGATTTTGATTGCTGGCGGT
CTAGATCGTGGCAATGAATTTGACGATTTGGTGCCAGACCTTTTAGGACTTAAGCAGATGATTATTTTGGG
AGAATCCGCAGAGCGTATGAAGCGAGCTGCTAACAAAGCAGAGGTCTCTTATCTTGAAGCTAGAAATGTGG
CAGAAGCAACAGAGCTTGCTTTTAAGCTGGCCCAAACAGGCGATACTATCTTGCTTAGCCCAGCCAATGCT
AGCTGGGATATGTATCCTAATTTTGAGGTTCGTGGGGATGAATTTTTGGCAACCTTTGATTGTTTAAGAGG
AGATGCCTAA

SEQ ID NO: 59 amino acid sequence comprising N-terminal leader sequence of GAS 366

MKVISNFQNKKILILGLAKSGEAAA

SEQ ID NO: 60 amino acid sequence comprising a fragment of GAS 366 where the N-terminal leader sequence is removed

KLLTKLGALVTVNDSKPFDQNPAAQALLEEGIKVICGSHPVELLDENFEYMVKNPGIPYDNPMVKRALAKE
IPILTEVELAYFVSEAPIIGITGSNGKTTTTTMIADVLNAGGQSALLSGNIGYPASKVVQKAIAGDTLVME
LSSFQLVGVNAFRPHIAVITNLMPTHLDYHGSFEDYVAAKWMIQAQMTESDYLILNANQEISATLAKTTKA
TVIPFSTQKVVDGAYLKDGILYFKEQAIIAATDLGVPGSHNIENALATIAVAKLSGIADDIIAQCLSHFGG
VKHRLQRVGQIKDITFYNDSKSTNILATQKALSGFDNSRLILIAGGLDRGNEFDDLVPDLLGLKQMIILGE
SAERMKRAANKAEVSYLEARNVAEATELAFKLAQTGDTILLSPANASWDMYPNFEVRGDEFLATFDCLRGD
A

SEQ ID NO: 61 amino acid sequence comprising GAS 159

MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSN
EAMYTKIKQGGTTYDIAVPSDYTIDKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPQNDYSLPYFWGTV
GIVYNDQLVDKAPMHWEDLWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNV
KAIVADEMKGYMIQGDAAIGITFSGEASEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFLNF
INRPENAAQNAAYIGYATPNKKAKALLPDEIKNDPAFYPTDDIIKKLEVYDNLGSRWLGIYNDLYLQFKMY
RK

SEQUENCE LISTING

SEQ ID NO: 62 polynucleotide sequence encoding GAS 159

ATGCGTAAACTTTATTCCTTTCTAGCAGGAGTTTTGGGTGTTATTGTTATTTTAACAAGTCTTTCTTTCAT
CTTGCAGAAAAAATCGGGTTCTGGTAGTCAATCGGATAAATTAGTTATTTATAACTGGGGAGATTACATTG
ATCCAGCTTTGCTCAAAAAATTCACCAAAGAAACGGGCATTGAAGTGCAGTATGAAACTTTCGATTCCAAT
GAAGCCATGTACACTAAAATCAAGCAGGGCGGAACCACTTACGACATTGCTGTTCCTAGTGATTACACCAT
TGATAAAATGATCAAAGAAAACCTACTCAATAAGCTTGATAAGTCAAAATTAGTTGGCATGGATAATATCG
GGAAAGAATTTTTAGGGAAAAGCTTTGACCCACAAAACGACTATTCTTTGCCTTATTTCTGGGGAACCGTT
GGGATTGTTTATAATGATCAATTAGTTGATAAGGCGCCTATGCACTGGGAAGATCTGTGGCGTCCAGAATA
TAAAAATAGTATTATGCTGATTGATGGAGCGCGTGAAATGCTAGGGGTTGGTTTAACAACTTTTGGTTATA
GTGTGAATTCTAAAAATCTAGAGCAGTTGCAGGCAGCCGAGAGAAAACTGCAGCAGTTGACGCCGAATGTT
AAAGCCATTGTAGCAGATGAGATGAAAGGCTACATGATTCAAGGTGACGCTGCTATTGGAATTACCTTTTC
TGGTGAAGCCAGTGAGATGTTAGATAGTAACGAACACCTTCACTACATCGTGCCTTCAGAAGGGTCTAACC
TTTGGTTTGATAATTTGGTACTACCAAAAACCATGAAACACGAAAAAGAAGCTTATGCTTTTTTGAACTTT
ATCAATCGTCCTGAAAATGCTGCGCAAAATGCTGCATATATTGGTTATGCGACACCAAATAAAAAAGCCAA
GGCCTTACTTCCAGATGAGATAAAAAATGATCCTGCTTTTTATCCAACAGATGACATTATCAAAAAATTGG
AAGTTTATGACAATTTAGGGTCAAGATGGTTGGGGATTTATAATGATTTATACCTCCAATTTAAAATGTAT
CGCAAATAA

SEQ ID NO: 63 amino acid sequence comprising N-terminal leader sequence of GAS 159

MRKLYSFLAGVLGVIVILTSLSFI

SEQ ID NO: 64 amino acid sequence comprising a fragment of GAS 159 where the N-terminal leader sequence is removed

LQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSNEAMYTKIKQGGTTYDIAVPSDYTI
DKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPQNDYSLPYFWGTVGIVYNDQLVDKAPMHWEDLWRPEY
KNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNVKAIVADEMKGYMIQGDAAIGITFS
GEASEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFLNFINRPENAAQNAAYIGYATPNKKAK
ALLPDEIKNDPAFYPTDDIIKKLEVYDNLGSRWLGIYNDLYLQFKMYRK

SEQ ID NO: 65 amino acid sequence comprising C-terminal hydrophobic sequence of GAS 159

WLGIYNDLYLQFKMYRK

SEQ ID NO: 66 amino acid sequence comprising a fragment of GAS 159 where the C-terminal hydrophobic region is removed

MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSN
EAMYTKIKQGGTTYDIAVPSDYTIDKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPQNDYSLPYFWGTV
GIVYNDQLVDKAPMHWEDLWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNV
KAIVADEMKGYMIQGDAAIGITFSGEASEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFLNF
INRPENAAQNAAYIGYATPNKKAKALLPDEIKNDPAFYPTDDIIKKLEVYDNLGSR

SEQ ID NO: 67 amino acid sequence comprising a fragment of GAS 159 where the N-terminal leader sequence and the C-terminal hydrophobic region is removed

LQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSNEAMYTKIKQGGTTYDIAVPSDYTI
DKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPQNDYSLPYFWGTVGIVYNDQLVDKAPMHWEDLWRPEY
KNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNVKAIVADEMKGYMIQGDAAIGITFS
GEASEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFLNFINRPENAAQNAAYIGYATPNKKAK
ALLPDEIKNDPAFYPTDDIIKKLEVYDNLGSR

SEQ ID NO: 68 amino acid sequence comprising GAS 217

MAQRIIVITGASGGLAQAIVKQLPKEDSLILLGRNKERLEHCYQHIDNKECLELDITNPVAIEKMVAQIYQ
RYGRIDVLINNAGYGAFKGFEEFSAQEIADMFQVNTLASIHFACLIGQKMAEQGQGHLINIVSMAGLIASA
KSSIYSATKFALIGFSNALRLELADKGVYVTTVNPGPIATKFFDQADPSGHYLESVGKFTLQPNQVAKRLV
SIIGKNKRELNLPFSLAVTHQFYTLFPKLSDYLARKVFNYK

SEQ ID NO: 69 polynucleotide sequence encoding GAS 217

SEQUENCE LISTING

ATGGCACAAAGAATCATTGTTATCACGGGAGCTTCTGGAGGACTGGCTCAGGCAATTGTTAAGCAGTTACC
CAAGGAAGACAGCTTGATTTTATTAGGACGTAACAAAGAACGCCTAGAACACTGTTATCAGCATATTGACA
ACAAAGAATGCCTCGAGTTGGATATTACCAATCCAGTAGCCATTGAGAAAATGGTCGCCCAGATTTACCAG
CGCTATGGCCGTATTGATGTCTTGATTAATAATGCTGGCTACGGAGCTTTCAAAGGCTTTGAAGAGTTTTC
TGCCCAAGAAATAGCTGATATGTTTCAGGTTAACACCCTAGCGAGCATTCACTTTGCTTGCTTGATTGGTC
AGAAAATGGCAGAGCAGGGGCAAGGTCACCTTATTAATATTGTGTCCATGGCAGGCTTGATTGCGTCAGCC
AAATCGAGCATTTATTCAGCCACCAAGTTTGCCCTTATCGGATTTTCCAATGCCCTTCGCTTAGAATTAGC
GGATAAAGGGGTTTACGTGACCACCGTGAATCCAGGTCCCATTGCCACCAAGTTTTTTGACCAAGCTGACC
CGTCTGGACATTATTTGGAAAGCGTTGGTAAATTTACTCTCCAACCAAATCAAGTGGCTAAGCGTTTGGTT
TCTATTATCGGGAAAAATAAACGAGAATTGAATTTGCCCTTTAGTTTAGCGGTGACCCATCAATTTTACAC
CCTTTTCCCTAAATTATCTGATTATCTTGCAAGAAAGGTATTTAATTATAAATGA

SEQ ID NO: 70 amino acid sequence comprising GAS 309

MIEKYLESSIESKCQLIVLFFKTSYLPITEVAEKTGLTFLQLNHYCEELNAFFPGSLSMTIQKRMISCQFT
HPFKETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNSSAYRMREALIPLLRNFELKLSKNKI
VGEEYRIRYLIALLYSKFGIKVYDLTQQDKNTIHSFLSHSSTHLKTSPWLSESFSFYDILLALSWKRHQFS
VTIPQTRIFQQLKKLFVYDSLKKSSHDIIETYCQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIRQYCQ
LFEENDTFRLLLNPIITLLPNLKEQKASLVKALMFFSKSFLFNLQHFIPETNLFVSPYYKGNQKLYTSLKL
IVEEWMAKLPGKRDLNHKHFHLFCHYVEQSLRNIQPPLVVVFVASNFINAHLLTDSFPRYFSDKSIDFHSY
YLLQDNVYQIPDLKPDLVITHSQLIPFVHHELTKGIAVAEISFDESILSIQELMYQVKEEKFQADLTKQLT

SEQ ID NO: 71 polynucleotide sequence encoding GAS 309

TTGATAGAAAAATACTTGGAATCATCAATCGAATCAAAATGTCAGTTAATTGTCTTGTTTTTTAAGACATC
TTATTTGCCAATAACTGAGGTAGCAGAAAAAACTGGCTTAACCTTTTTACAACTAAACCATTATTGTGAGG
AACTGAATGCCTTTTTCCCTGGTAGTCTGTCTATGACCATCCAAAAAAGGATGATATCTTGCCAATTTACA
CATCCTTTTAAAGAAACTTATCTTTACCAACTCTATGCATCATCTAATGTCTTACAATTACTAGCCTTTTT
AATAAAAAATGGTTCCCACTCTCGTCCCCTTACGGATTTTGCAAGAAGTCATTTTTTATCAAACTCCTCAG
CTTATCGGATGCGCGAAGCATTGATTCCTTTATTAAGAAACTTTGAATTAAAACTCTCTAAGAACAAGATT
GTCGGTGAGGAATATCGCATCCGTTACCTCATCGCTCTGCTATATAGTAAGTTTGGCATTAAAGTTTATGA
CTTGACGCAGCAAGACAAAAACACTATTCATAGCTTTTTATCCCATAGTTCCACCCACCTTAAAACCTCTC
CTTGGTTATCGGAATCGTTTTCTTTCTATGACATTTTATTAGCTTTATCGTGGAAGCGGCATCAATTTTCG
GTAACTATTCCCCAAACCAGAATTTTTCAACAATTAAAAAAACTTTTTGTCTACGATTCTTTGAAAAAAAG
TAGCCATGATATTATCGAAACTTACTGCCAACTAAACTTTTCAGCAGGAGATTTGGACTACCTCTATTTAA
TTTATATCACCGCTAATAATTCTTTTGCGAGCTTACAATGGACACCTGAGCATATCAGACAATATTGTCAA
CTTTTTGAAGAAAATGATACTTTTCGCCTGCTTTTAAATCCTATCATCACTCTTTTACCTAACCTAAAAGA
GCAAAAGGCTAGTTTAGTAAAAGCTCTTATGTTTTTTTCAAAATCATTCTTGTTTAATCTGCAACATTTTA
TTCCTGAGACCAACTTATTCGTTTCTCCGTACTATAAAGGAAACCAAAAACTCTATACGTCCTTAAAGTTA
ATTGTCGAAGAGTGGATGGCCAAACTTCCTGGTAAGCGTGACTTGAACCATAAGCATTTTCATCTTTTTTG
CCACTATGTCGAGCAAAGTCTAAGAAATATCCAACCTCCTTTAGTTGTTGTTTTCGTAGCCAGTAATTTTA
TCAATGCTCATCTCCTAACGGATTCTTTTCCAAGGTATTTCTCGGATAAAAGCATTGATTTTCATTCCTAT
TATCTATTGCAAGATAATGTTTATCAAATTCCTGATTTAAAGCCAGATTTGGTCATCACTCACAGTCAACT
GATTCCTTTTGTTCACCATGAACTTACAAAAGGAATTGCTGTTGCTGAAATATCTTTTGATGAATCGATTC
TGTCTATCCAAGAATTGATGTATCAAGTTAAAGAGGAAAAATTCCAAGCTGATTTAACCAAGCAATTAACA
TAA

SEQ ID NO: 72 amino acid sequence comprising GAS 372

MIQIGKLFAGRYRILKSIGRGGMADVYLANDLILDNEDVAIKVLRTNYQTDQVAVARFQREARAMAELNHP
NIVAIRDIGEEDGQQFLVMEYVDGADLKRYIQNHAPLSNNEVVRIMEEVLSAMTLAHQKGIVHRDLKPQNI
LLTKEGVVKVTDFGIAVAFAETSLTQTNSMLGSVHYLSPEQARGSKATIQSDIYAMGIMLFEMLTGHIPYD
GDSAVTIALQHFQKPLPSIIEENHNVPQALENVVIRATAKKLSDRYGSTFEMSRDLMTALSYNRSRERKII
FENVESTKPLPKVASGPTASVKLSPPTPTVLTQESRLDQTNQTDALQPPTKKKKSGRFLGTLFKILFSFFI
VGVALFTYLILTKPTSVKVPNVAGTSLKVAKQELYDVGLKVGKIRQIESDTVAEGNVVRTDPKAGTAKRQG
SSITLYVSIGNKGFDMENYKGLDYQEAMNSLIETYGVPKSKIKIERIVTNEYPENTVISQSPSAGDKFNPN
GKSKITLSVAVSDTITMPMVTEYSYADAVNTLTALGIDASRIKAYVPSSSSATGFVPIHSPSSKAIVSGQS
PYYGTSLSLSDKGEISLYLYPEETHSSSSSSSSTSSSNSSSINDSTAPGSNTELSPSETTSQTP

SEQUENCE LISTING

SEQ ID NO: 73 polynucleotide sequence encoding GAS 372

ATGATTCAGATTGGCAAATTATTTGCTGGTCGTTATCGCATTCTGAAATCTATTGGCCGCGGTGGTATGGC
GGATGTTTATTTAGCAAATGACTTGATCTTGGATAATGAAGACGTTGCAATCAAGGTCTTGCGTACCAATT
ATCAAACAGATCAGGTAGCAGTTGCGCGTTTCCAACGAGAAGCGCGGGCCATGGCTGAATTGAACCATCCC
AATATTGTTGCCATCCGGGATATAGGTGAAGAAGACGGACAGCAATTTTTAGTAATGGAATATGTGGATGG
TGCTGACCTAAAGAGATACATTCAAAATCATGCTCCATTATCTAATAATGAAGTGGTTAGAATTATGGAAG
AAGTCCTTTCTGCTATGACTTTAGCCCACCAAAAAGGAATTGTACACAGAGATTTAAAACCTCAAAATATC
CTACTAACTAAGGAGGGTGTTGTCAAAGTAACTGATTTCGGCATCGCAGTAGCCTTTGCAGAAACAAGCTT
GACACAAACTAATTCGATGTTAGGCAGTGTTCATTACTTGTCTCCAGAACAGGCTCGCGGCTCCAAAGCGA
CGATTCAAAGTGATATTTATGCGATGGGGATTATGCTCTTTGAGATGTTGACAGGCCATATCCCTTATGAC
GGCGATAGTGCTGTTACGATTGCCTTGCAACATTTTCAAAAGCCTCTTCCATCTATTATCGAGGAGAACCA
CAATGTGCCACAAGCTTTGGAGAATGTTGTTATTCGAGCAACAGCCAAGAAATTAAGTGATCGTTACGGGT
CAACCTTTGAAATGAGTCGTGACTTAATGACGGCGCTTAGTTATAATCGTAGTCGGGAGCGTAAGATTATC
TTTGAGAATGTTGAAAGTACCAAACCCCTCCCCAAAGTGGCCTCAGGTCCCACCGCTTCTGTAAAATTGTC
TCCCCCTACCCCAACAGTGTTAACACAGGAAAGTCGATTAGATCAAACTAATCAAACAGATGCTTTACAGC
CCCCCACCAAAAAGAAAAAAAGTGGTCGTTTTTTAGGTACTTTATTCAAAATTCTTTTTTCTTTCTTTATT
GTAGGTGTAGCACTCTTTACTTATCTTATACTAACTAAACCAACTTCTGTGAAAGTTCCTAATGTAGCAGG
CACTAGTCTTAAAGTTGCCAAACAAGAACTGTATGATGTTGGGCTAAAAGTGGGTAAAATCAGGCAAATTG
AGAGTGATACGGTTGCTGAGGGAAATGTAGTTAGAACAGATCCTAAAGCAGGAACAGCTAAGAGGCAAGGC
TCAAGCATTACGCTTTATGTGTCAATTGGAAACAAAGGTTTTGACATGGAAAACTACAAAGGACTAGATTA
TCAAGAAGCTATGAATAGTTTGATAGAAACTTATGGTGTTCCAAAATCAAAAATCAAAATTGAGCGCATTG
TAACTAATGAATATCCTGAAAATACAGTCATCAGTCAATCGCCAAGTGCGGGTGATAAATTTAATCCAAAC
GGAAAGTCTAAAATTACGCTCAGTGTTGCTGTTAGTGATACGATCACTATGCCTATGGTAACAGAATATAG
TTATGCAGATGCAGTCAATACCTTAACAGCTTTAGGTATAGATGCATCTAGAATAAAAGCTTATGTGCCAA
GCTCTAGCTCAGCAACGGGCTTTGTGCCAATTCATTCTCCTAGTTCTAAAGCTATTGTCAGTGGTCAATCT
CCTTACTATGGAACGTCTTTGAGTCTGTCTGATAAAGGAGAGATTAGTCTTTACCTTTATCCAGAAGAAAC
ACACTCTTCTAGTAGCTCATCGAGTTCAACGTCAAGTTCAAACAGTTCTTCAATAAATGATAGTACTGCAC
CAGGTAGCAACACTGAATTAAGCCCATCAGAAACTACTTCTCAAACACCTTAA

SEQ ID NO: 74 amino acid sequence comprising GAS 39

MDLILFLLVLVLLGLGAYLLFKVNGLQHQLAQTLEGNADNLSDQMTYQLDTANKQQLLELTQLMNRQQAGL
YQQLTDIRDVLHRSLSDSRDRSDKRLEKINQQVNQSLKNMQESNEKRLEKMRQIVEEKLEETLKNRLHASF
DSVSKQLESVNKGLGEMRSVAQDVGTLNKVLSNTKTRGILGELQLGQIIEDIMTSSQYEREFVTVSGSSER
VEYAIKLPGNGQGGYIYLPIDSKFPLEDYYRLEDAYEVGDKLAIEASRKALLAAIKRFAKDIHKKYLNPPE
TTNFGVMFLPTEGLYSEVVRNASFFDSLRREENIVVAGPSTLSALLNSLSVGFKTLNIQKNADDISKILGN
VKLEFDKFGGLLAKAQKQMNTANNTLDQLISTRTNAIVRALNTVETYQDQATKSLLNMPLLEEENNEN

SEQ ID NO: 75 polynucleotide sequence encoding GAS 39

ATGGACCTTATCTTGTTCCTTTTGGTCTTGGTTCTCTTAGGTTTAGGGGCTTATCTGTTGTTCAAAGTCAA
CGGCCTTCAACATCAGCTTGCCCAAACCCTAGAAGGCAACGCGGATAATTTGTCTGACCAAATGACCTACC
AGTTGGATACAGCTAACAAACAACAATTGTTAGAGCTAACACAGCTGATGAACCGACAACAAGCAGGCCTT
TACCAACAATTAACAGATATTCGTGACGTCTTGCACCGTAGTTTGTCTGATAGTAGGGACCGGTCTGACAA
ACGCTTAGAAAAAATTAACCAGCAGGTCAACCAATCGCTCAAAAATATGCAAGAATCTAACGAAAAACGTT
TGGAGAAAATGCGCCAGATCGTTGAAGAAAAATTGGAAGAAACCTTAAAAAATCGTCTGCACGCCTCTTTC
GATTCTGTATCCAAGCAACTAGAAAGTGTCAATAAAGGCTTGGGAGAAATGCGTAGCGTGGCTCAAGATGT
GGGTACTTTAAATAAGGTTTTGTCCAATACCAAAACACGAGGCATTTTAGGCGAACTTCAACTAGGCCAAA
TCATTGAGGATATCATGACATCAAGCCAGTACGAAAGAGAATTTGTAACGGTTAGTGGTTCTAGTGAACGC
GTAGAATATGCGATTAAGCTCCCAGGAAATGGTCAAGGCGGTTATATTTACCTACCGATTGACTCAAAATT
CCCTCTTGAAGATTATTACCGATTAGAAGATGCTTACGAAGTTGGTGATAAACTGGCCATCGAGGCTAGCC
GAAAAGCACTTCTGGCAGCTATCAAACGCTTTGCCAAAGACATTCATAAAAAGTACTTGAACCCCCCAGAG
ACGACCAATTTCGGAGTTATGTTCTTACCAACAGAAGGTCTTTATTCAGAAGTGGTCAGAAATGCGTCTTT
CTTTGATAGCCTTCGTCGGGAAGAAAATATTGTGGTTGCAGGCCCTTCGACCCTGTCTGCTTTGCTGAATT
CCTTATCTGTTGGTTTCAAGACCCTTAATATCCAAAAAAATGCTGATGACATCAGTAAAATTTTAGGCAAT
GTCAAGTTAGAATTCGATAAATTTGGCGGCCTGCTTGCCAAGGCTCAAAAACAAATGAATACAGCTAATAA
TACGCTGGATCAGCTCATTTCAACAAGGACAAATGCCATTGTTCGAGCCTTGAATACCGTTGAAACTTATC
AAGACCAAGCAACAAAATCTCTCTTGAACATGCCCTTATTAGAAGAGGAAAATAATGAAAATTAA

SEQUENCE LISTING

SEQ ID NO: 76 amino acid sequence comprising GAS 42

MTKEKLVAFSQAHAEPAWLQERRLAALEAIPNLELPTIERVKFHRWNLGDGTLTENESLASVPDFIAIGDN
PKLVQVGTQTVLEQLPMALIDKGVVFSDFYTALEEIPEVIEAHFGQALAFDEDKLAAYHTAYFNSAAVLYV
PDHLEITTPIEAIFLQDSDSDVPFNKHVLVIAGKESKFTYLERFESIGNATQKISANISVEVIAQAGSQIK
FSAIDRLGPSVTTYISRRGRLEKDANIDWALAVMNEGNVIADFDSDLIGQGSQADLKVVAASSGRQVQGID
TRVTNYGQRTVGHILQHGVILERGTLTFNGIGHILKDAKGADAQQESRVLMLSDQARADANPILLIDENEV
TAGHAASIGQVDPEDMYYLMSRGLDQETAERLVIRGFLGAVIAEIPIPSVRQEIIKVLDEKLLNR

SEQ ID NO: 77 polynucleotide sequence encoding GAS 42

ATGACAAAAGAAAAACTAGTGGCTTTTTCGCAAGCCCACGCTGAGCCTGCTTGGCTGCAAGAACGGCGTTT
AGCGGCATTAGAAGCCATTCCAAATTTGGAATTACCAACCATCGAAAGGGTTAAATTTCACCGTTGGAATC
TAGGAGATGGTACCTTAACAGAAAATGAAAGTCTAGCTAGTGTTCCAGATTTTATAGCTATTGGAGATAAC
CCAAAGCTTGTTCAGGTAGGCACGCAAACAGTCTTAGAACAGTTACCAATGGCGTTAATTGACAAGGGAGT
TGTTTTCAGTGATTTTTATACGGCGCTTGAGGAAATCCCAGAAGTAATTGAAGCTCATTTTGGTCAGGCAT
TAGCTTTTGATGAAGACAAACTAGCTGCCTACCACACTGCTTATTTTAATAGCGCAGCCGTGCTCTACGTT
CCTGATCACTTGGAAATCACAACTCCTATTGAAGCTATTTTCTTACAAGATAGTGACAGTGACGTTCCTTT
TAACAAGCATGTTCTAGTGATTGCAGGAAAAGAAAGTAAGTTCACCTATTTAGAGCGTTTTGAATCTATTG
GCAATGCCACTCAAAAGATCAGCGCTAATATCAGTGTAGAAGTGATTGCTCAAGCAGGCAGCCAGATTAAA
TTCTCGGCTATCGACCGCTTAGGTCCTTCAGTGACAACCTATATTAGCCGTCGAGGACGTTTAGAGAAGGA
TGCCAACATTGATTGGGCCTTAGCTGTGATGAATGAAGGCAATGTCATTGCTGATTTTGACAGTGATTTGA
TTGGTCAGGGCTCACAAGCTGATTTGAAAGTTGTTGCAGCCTCAAGTGGTCGTCAGGTACAAGGTATTGAC
ACGCGCGTGACCAACTATGGTCAACGTACGGTCGGTCATATTTTACAGCATGGTGTGATTTTGGAACGTGG
CACCTTAACGTTTAACGGGATTGGTCATATTCTAAAAGACGCTAAGGGAGCTGATGCTCAACAAGAAAGCC
GTGTTTTGATGCTTTCTGACCAAGCAAGAGCCGATGCCAATCCAATCCTCTTAATTGATGAAAATGAAGTA
ACAGCAGGTCATGCAGCTTCTATCGGTCAGGTTGACCCTGAAGATATGTATTACTTGATGAGTCGAGGACT
GGATCAAGAAACAGCAGAACGATTGGTTATTAGAGGATTCCTAGGAGCGGTTATCGCTGAAATTCCTATTC
CATCAGTCCGCCAAGAGATTATTAAGGTTTTAGATGAGAAATTGCTTAATCGTTAA

SEQ ID NO: 78 amino acid sequence comprising GAS 58

MKWSGFMKTKSKRFLNLATLCLALLGTTLLMAHPVQAEVISKRDYMTRFGLGDLEDDSANYPSNLEARYKG
YLEGYEKGLKGDDIPERPKIQVPEDVQPSDHGDYRDGYEEGFGEGQHKRDPLETEAEDDSQGGRQEGRQGH
QEGADSSDLNVEESDGLSVIDEVVGVIYQAFSTIWTYLSGLF

SEQ ID NO: 79 polynucleotide sequence encoding GAS 58

ATGAAATGGAGTGGTTTTATGAAAACAAAATCAAAACGCTTTTTAAACCTAGCAACCCTTTGCTTGGCCCT
ACTAGGAACAACTTTGCTAATGGCACATCCCGTACAGGCGGAGGTGATATCAAAAAGAGACTATATGACTC
GCTTCGGGTTAGGCGATTTAGAAGATGATTCAGCTAACTATCCTTCAAATTTAGAAGCTAGATATAAAGGA
TATTTAGAGGGATATGAAAAAGGCTTAAAAGGAGATGATATACCCGAACGGCCCAAGATTCAGGTTCCTGA
GGATGTTCAGCCATCTGACCATGGCGACTATAGAGATGGTTATGAGGAAGGATTTGGAGAAGGACAACATA
AACGTGATCCATTAGAAACAGAAGCAGAAGATGATTCTCAAGGAGGACGTCAAGAAGGACGTCAAGGACAT
CAAGAAGGAGCAGATTCTAGTGATTTGAACGTTGAAGAAAGCGACGGTTTGTCTGTTATTGATGAAGTAGT
TGGAGTAATTTATCAAGCATTTAGTACTATTTGGACATACTTAAGCGGTTTGTTCTAA

SEQ ID NO: 80 amino acid sequence comprising N-terminal leader sequence of GAS 58

MKWSGFMKTKSKRFLNLATLCLALLGTTLLMA

SEQ ID NO: 81 amino acid sequence comprising a fragment of GAS 58 where the N-terminal leader sequence is removed

HPVQAEVISKRDYMTRFGLGDLEDDSANYPSNLEARYKGYLEGYEKGLKGDDIPERPKIQVPEDVQPSDHG
DYRDGYEEGFGEGQHKRDPLETEAEDDSQGGRQEGRQGHQEGADSSDLNVEESDGLSVIDEVVGVIYQAFS
TIWTYLSGLF

SEQ ID NO: 82 amino acid sequence comprising GAS 290

SEQUENCE LISTING

MKHILFIVGSLREGSFNHQLAAQAQKALEHQAVVSYLNWKDVPVLNQDIEANAPLPVVDARQAVQSADAIW
IFTPVYNFSIPGSVKNLLDWLSRALDLSDPTGPSAIGGKVVTVSSVANGGHDQVFDQFKALLPFIRTSVAG
EFTKATVNPDAWGTGRLEISKETKANLLSQAEALLAAI

SEQ ID NO: 83 polynucleotide sequence encoding GAS 290

ATGAAACATATTTTATTTATTGTTGGCTCGCTTCGTGAAGGGTCTTTTAACCATCAATTAGCGGCTCAAGC
ACAAAAAGCTCTGGAACATCAAGCAGTTGTATCTTACTTAAATTGGAAAGACGTTCCTGTTTTGAATCAAG
ATATCGAAGCTAATGCACCTTTACCAGTTGTTGACGCTCGTCAAGCTGTTCAGTCAGCGGATGCTATCTGG
ATTTTTACACCAGTTTACAACTTCTCTATTCCAGGTTCTGTTAAAAACCTGCTAGACTGGTTGTCTCGTGC
TCTTGATTTGTCTGATCCGACGGGCCCATCTGCTATTGGCGGTAAGGTGGTTACGGTCTCTTCAGTTGCAA
ATGGCGGGCATGATCAAGTATTTGATCAGTTTAAAGCACTATTGCCGTTTATCCGAACTTCAGTAGCAGGA
GAGTTTACAAAAGCAACTGTGAATCCTGATGCCTGGGGAACAGGAAGGCTTGAGATTTCAAAAGAGACAAA
AGCAAACTTGCTATCTCAGGCAGAGGCTCTTTTAGCGGCTATTTAG

SEQ ID NO: 84 amino acid sequence comprising GAS 511

MTDVSRILKEARDQGRLTTLDYANLIFDDFMELHGDRHFSDDGAIVGGLAYLAGQPVTVIGIQKGKNLQDN
LARNFGQPNPEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQGEAIAKNLMEMSDLKVPIIAI
IIGEGGSGGALALAVADQVWMLENTMYAVLSPEGFASILWKDGSRATEAAELMKITAGELYKMGIVDRIIP
EHGYFSSEIVDIIKANLIEQITSLQAKPLDQLLDERYQRFRKY

SEQ ID NO: 85 polynucleotide sequence encoding GAS 511

ATGACAGATGTATCAAGAATTTTAAAAGAAGCGCGTGATCAAGGGCGTTTAACAACTTTGGATTACGCCAA
CCTTATTTTCGATGACTTTATGGAACTGCATGGCGATCGCCATTTTTCAGATGATGGTGCCATTGTAGGTG
GCCTAGCTTATTTGGCGGGACAACCTGTTACGGTCATTGGTATTCAAAAAGGTAAGAATTTACAGGATAAT
TTGGCAAGGAATTTTGGCCAGCCCAATCCAGAAGGTTATCGTAAAGCTTTGCGCCTTATGAAACAGGCAGA
AAAATTTGGACGACCAGTTGTTACGTTTATCAATACTGCAGGAGCCTATCCAGGTGTCGGTGCGGAAGAAC
GAGGACAGGGTGAGGCCATTGCTAAAAATTTGATGGAAATGAGTGATCTCAAGGTTCCCATTATCGCCATC
ATTATTGGTGAAGGAGGCTCTGGTGGTGCATTAGCCTTAGCGGTTGCCGATCAGGTCTGGATGCTTGAAAA
TACTATGTATGCGGTTCTTAGCCCAGAAGGCTTTGCTTCTATTTTATGGAAGGATGGTTCAAGGGCGACCG
AGGCCGCTGAATTGATGAAAATCACAGCGGGTGAACTCTACAAAATGGGAATAGTAGACCGTATTATTCCA
GAACATGGTTATTTTTCAAGTGAAATCGTTGACATCATCAAAGCTAACCTCATCGAACAAATAACCAGTTT
GCAAGCTAAGCCATTAGACCAATTATTAGATGAGCGCTACCAACGCTTTCGTAAATATTAA

SEQ ID NO: 86 amino acid sequence comprising GAS 533

MAITVADIRREVKEKNVTFLRLMFTDIMGVMKNVEIPATKEQLDKVLSNKVMFDGSSIEGFVRINESDMYL
YPDLDTWIVFPWGDENGAVAGLICDIYTAEGKPFAGDPRGNLKRALKHMNEIGYKSFNLGPEPEFFLFKMD
DKGNPTLEVNDNGGYFDLAPIDLADNTRREIVNILTKMGFEVEASHHEVAVGQHEIDFKYADVLKACDNIQ
IFKLVVKTIAREHGLYATFMAKPKFGIAGSGMHCNMSLFDNQGNNAFYDEADKRGMQLSEDAYYFLGGLMK
HAYNYTAITNPTVNSYKRLVPGYEAPVYVAWAGSNRSPLIRVPASRGMGTRLELRSVDPTANPYLALAVLL
EAGLDGIINKIEAPEPVEANIYTMTMEERNEAGIIDLPSTLHNALKALQKDDVVQKALGYHIYTNFLEAKR
IEWSSYATFVSQWEIDHYIHNY

SEQ ID NO: 87 polynucleotide sequence encoding GAS 533

ATGGCAATAACAGTAGCTGACATTCGTCGTGAAGTCAAAGAAAAAAATGTAACGTTTCTTCGCTTGATGTT
CACTGATATCATGGGCGTTATGAAAAATGTGGAGATTCCTGCAACTAAAGAACAGTTAGACAAAGTATTGT
CTAACAAGGTTATGTTTGATGGTTCATCTATCGAAGGTTTTGTACGGATCAATGAGTCAGATATGTACCTT
TACCCCGATTTAGACACTTGGATTGTTTTTCCCTGGGGAGATGAAAATGGAGCAGTTGCAGGTTTAATTTG
TGATATTTATACAGCAGAAGGAAAGCCTTTTGCAGGAGATCCTAGAGGAAATTTAAAAAGAGCCCTGAAAC
ACATGAACGAGATCGGCTACAAATCATTTAATCTTGGACCAGAACCAGAATTTTTCCTTTTTAAGATGGAT
GATAAAGGTAATCCGACACTTGAAGTTAACGATAATGGTGGTTATTTTGATTTAGCGCCAATTGACTTAGC
AGACAACACGCGCCGTGAAATTGTGAATATTTTAACGAAAATGGGTTTTGAAGTGGAAGCTAGTCATCATG
AAGTGGCTGTTGGTCAACATGAGATTGATTTTAAATATGCAGATGTTTTGAAAGCTTGTGATAATATTCAA
ATTTTTAAGCTAGTTGTAAAAACGATTGCCCGTGAACATGGACTTTATGCTACTTTCATGGCTAAACCAAA
ATTTGGAATAGCTGGATCAGGGATGCACTGTAACATGTCTTTGTTTGATAACCAAGGTAATAATGCTTTTT
ATGATGAAGCTGATAAGCGAGGGATGCAGTTATCAGAAGATGCTTATTATTTCTTGGGAGGACTAATGAAG
CATGCTTATAACTACACTGCTATCACTAACCCTACAGTGAATTCTTATAAACGATTAGTTCCAGGTTATGA

SEQUENCE LISTING

GGCACCTGTTTATGTCGCTTGGGCTGGAAGTAATCGTTCACCGCTTATCCGTGTTCCAGCATCACGTGGTA
TGGGAACGCGTTTGGAGTTACGTTCGGTTGATCCGACAGCTAATCCTTATTTAGCCTTGGCTGTTCTCTTG
GAAGCTGGATTAGATGGTATCATTAACAAAATTGAAGCTCCAGAACCCGTTGAAGCTAACATTTATACCAT
GACAATGGAAGAACGAAATGAAGCAGGCATTATTGATTTGCCATCAACGCTTCATAATGCCTTAAAAGCTC
TTCAAAAAGATGATGTGGTACAAAAGGCACTAGGTTACCATATCTACACTAATTTCTTAGAAGCAAAACGA
ATTGAATGGTCTTCCTATGCAACTTTTGTTTCTCAATGGGAAATTGACCATTATATTCATAATTATTAG

SEQ ID NO: 88 amino acid sequence comprising GAS 527

MTEISILNDVQKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITAQELREINPIGIVLSGGPNSVYADNA
FGIDPEIFELGIPILGICYGMQLITHKLGGKVVPAGQAGNREYGQSTLHLRETSKLFSGTPQEQLVLMSHG
DAVTEIPEGFHLVGDSNDCPYAAIENTEKNLYGIQFHPEVRHSVYGNDILKNFAISICGARGDWSMDNFID
MEIAKIRETVGDRKVLLGLSGGVDSSVVGVLLQKAIGDQLTCIFVDHGLLRKDEGDQVMGMLGGKFGLNII
RVDASKRFLDLLADVEDPEKKRKIIGNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHN
VGGLPEDMQFELIEPLNTLFKDEVRALGIALGMPEEIVWRQPFPGPGLAIRVMGAITEEKLETVRESDAIL
REEIAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADFAQLPWDVLKKISTRIVN
EVDHVNRIVYDITSKPPATVEWE

SEQ ID NO: 89 polynucleotide sequence encoding GAS 527

ATGACTGAAATTTCAATTTTGAATGATGTTCAAAAAATTATCGTTCTTGATTATGGTAGCCAGTACAATCA
GCTTATTGCTAGACGTATTCGAGAGTTTGGTGTTTTCTCCGAACTAAAAAGCCATAAAATCACCGCTCAAG
AACTTCGTGAGATCAATCCCATAGGTATCGTTTTATCAGGAGGGCCTAACTCTGTTTACGCTGATAACGCC
TTTGGCATTGACCCTGAAATCTTTGAACTAGGGATTCCGATTCTTGGTATCTGTTACGGTATGCAATTAAT
CACCCATAAATTAGGTGGTAAAGTTGTTCCTGCTGGACAAGCTGGTAATCGTGAATACGGTCAGTCAACCC
TTCATCTTCGTGAAACGTCAAAATTATTTTCAGGCACACCTCAAGAACAACTCGTTTTGATGAGCCATGGT
GATGCTGTTACTGAAATTCCAGAAGGTTTCCACCTTGTTGGAGACTCAAATGACTGTCCCTATGCAGCTAT
TGAAAATACTGAGAAAAACCTTTACGGTATTCAGTTCCACCCAGAAGTGAGACACTCTGTTTATGGAAATG
ACATTCTTAAAAACTTTGCTATATCAATTTGTGGCGCGCGTGGTGATTGGTCAATGGATAATTTTATTGAC
ATGGAAATTGCTAAAATTCGTGAAACTGTAGGCGATCGTAAAGTTCTTCTAGGTCTTTCTGGTGGAGTTGA
TTCTTCAGTTGTTGGTGTTCTACTTCAAAAAGCTATCGGTGACCAATTAACTTGTATTTTCGTTGATCACG
GTCTTCTTCGTAAAGACGAGGGCGATCAAGTTATGGGAATGCTTGGGGGCAAATTTGGCCTAAATATTATC
CGTGTGGATGCTTCAAAACGTTTCTTAGACCTTCTTGCAGACGTTGAAGATCCTGAGAAAAAACGTAAAAT
TATTGGTAATGAATTTGTCTATGTTTTTGATGATGAAGCCAGCAAATTAAAAGGTGTTGACTTCCTTGCCC
AAGGAACACTTTATACTGATATCATTGAGTCAGGAACAGAAACTGCTCAAACCATCAAATCACATCACAAT
GTGGGTGGTCTCCCCGAAGACATGCAGTTTGAATTGATTGAGCCCTTAAACACTCTTTTCAAAGATGAAGT
TCGAGCGCTTGGAATCGCTCTTGGAATGCCTGAAGAAATTGTTTGGCGCCAACCATTTCCAGGTCCTGGAC
TTGCTATCCGTGTCATGGGAGCAATTACTGAAGAAAAACTTGAAACCGTTCGCGAATCAGACGCTATCCTT
CGTGAAGAAATTGCTAAGGCTGGACTTGATCGTGACGTGTGGCAATACTTTACAGTTAACACAGGTGTCCG
TTCTGTAGGCGTCATGGGAGATGGTCGTACTTATGATTATACCATCGCCATTCGTGCTATTACGTCTATTG
ATGGTATGACAGCTGACTTTGCTCAACTTCCTTGGGATGTCTTGAAAAAAATCTCAACACGTATCGTAAAT
GAAGTTGACCACGTTAACCGTATCGTCTACGACATCACAAGTAAACCACCCGCAACAGTTGAATGGGAATA
A

SEQ ID NO: 90 amino acid sequence comprising GAS 294

MSQSTATYINVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVKATPQHKTTNFAELVCSNSFRGDSLTNAVGL
LKEEMRRLDSIIMRNGEANRVPAGGAMAVDREGYAESVTAELENHPLIEVIRGEITEIPDDAITVIATGPL
TSDALAEKIHALNGGDGFYFYDAAAPIIDKSTIDMSKVYLKSRYDKGEAAYLNCPMTKEEFMAFHEALTTA
EEAPLNAFEKEKYFEGCMPIEVMAKRGIKTMLYGPMKPVGLEYPDDYTGPRDGEFKTPYAVVQLRQDNAAG
SLYNIVGFQTHLKWGEQKRVFQMIPGLENAEFVRYGVMHRNSYMDSPNLLTETFQSRSNPNLFFAGQMTGV
EGYVESAASGLVAGINAARLFKREEALIFPQTTAIGSLPHYVTHADSKHFQPMNVNFGIIKELEGPRIRDK
KERYEAIASRALADLDTCLASL

SEQ ID NO: 91 polynucleotide sequence encoding GAS 294

TTGTCTCAATCAACTGCAACTTATATTAATGTTATTGGAGCTGGGCTAGCTGGTTCTGAAGCTGCCTATCA
GATTGCTAAGCGCGGTATCCCCGTTAAATTGTATGAAATGCGTGGTGTCAAAGCAACACCGCAACATAAAA
CCACTAATTTTGCCGAATTGGTCTGTTCCAACTCATTTCGTGGTGATAGCTTAACCAATGCAGTCGGTCTT
CTCAAAGAAGAAATGCGGCGATTAGACTCCATTATTATGCGTAATGGTGAAGCTAACCGCGTACCTGCTGG

SEQUENCE LISTING

GGGAGCAATGGCTGTTGACCGTGAGGGGTATGCAGAGAGTGTCACTGCAGAGTTGGAAAATCATCCTCTCA
TTGAGGTCATTCGTGGTGAAATTACAGAAATCCCTGACGATGCTATCACGGTTATCGCGACGGGACCGCTG
ACTTCGGATGCCCTGGCAGAAAAAATTCACGCGCTAAATGGTGGCGACGGATTCTATTTTTACGATGCAGC
AGCGCCTATCATTGATAAATCTACCATTGATATGAGCAAGGTTTACCTTAAATCTCGCTACGATAAAGGCG
AAGCTGCTTACCTCAACTGCCCTATGACCAAAGAAGAATTCATGGCTTTCCATGAAGCTCTGACAACCGCA
GAAGAAGCCCCGCTGAATGCCTTTGAAAAAGAAAAGTATTTTGAAGGCTGTATGCCGATTGAAGTTATGGC
TAAACGTGGCATTAAAACCATGCTTTATGGACCTATGAAACCCGTTGGATTGGAATATCCAGATGACTATA
CAGGTCCTCGCGATGGAGAATTTAAAACGCCATATGCCGTCGTGCAATTGCGTCAAGATAATGCAGCTGGA
AGCCTTTATAATATCGTTGGTTTCCAAACCCATCTCAAATGGGGTGAGCAAAAACGCGTTTTCCAAATGAT
TCCAGGGCTTGAAAATGCTGAGTTTGTCCGCTACGGCGTCATGCATCGCAATTCCTATATGGATTCACCAA
ATCTTTTAACCGAAACCTTCCAATCTCGGAGCAATCCAAACCTTTTCTTTGCAGGTCAGATGACTGGAGTT
GAAGGTTATGTCGAATCAGCTGCTTCAGGTTTAGTAGCAGGAATCAATGCTGCTCGTTTGTTCAAAAGAGA
AGAAGCACTTATTTTTCCTCAGACAACAGCTATTGGGAGTTTGCCTCATTATGTGACTCATGCCGACAGTA
AGCATTTCCAACCAATGAACGTCAACTTTGGCATCATCAAAGAGTTAGAAGGCCCACGCATTCGTGACAAA
AAAGAACGTTATGAAGCTATTGCTAGTCGTGCTTTGGCAGATTTAGACACCTGCTTAGCGTCGCTTTAA

SEQ ID NO: 92 amino acid sequence comprising GAS 253

MPKKILFTGGGTVGHVTLNLILIPKFIKDGWEVHYIGDKNGIEHTEIEKSGLDVTFHAIATGKLRRYFSWQ
NLADVFKVALGLLQSLFIVAKLRPQALFSKGGFVSVPPVVAAKLLGKPVFIHESDRSMGLANKIAYKFATT
MYTTFEQEDQLSKVKHLGAVTKVFKDANQMPESTQLEAVKEYFSRDLKTLLFIGGSAGAHVFNQFISDHPE
LKQRYNIINITGDPHLNELSSHLYRVDYVTDLYQPLMAMADLVVTRGGSNTLFELLAMAKLHLIVPLGKEA
SRGDQLENATYFEKRGYAKQLQEPDLTLHNFDQAMADLFEHQADYEATMLATKEIQSPDFFYDLLRADISS
AIKEK

SEQ ID NO: 93 polynucleotide sequence encoding GAS 253

ATGCCTAAGAAGATTTTATTTACAGGTGGTGGAACTGTAGGTCATGTCACCTTGAACCTCATTCTCATACC
AAAATTTATCAAGGACGGTTGGGAAGTACATTATATTGGTGATAAAAATGGCATTGAACATACAGAAATTG
AAAAGTCAGGCCTTGACGTGACCTTTCATGCTATCGCGACAGGCAAGCTTAGACGCTATTTTTCATGGCAA
AATCTAGCTGATGTTTTTAAGGTTGCACTTGGCCTCCTACAGTCTCTCTTTATTGTTGCCAAGCTTCGCCC
TCAAGCCCTTTTTTCCAAAGGTGGTTTTGTCTCAGTACCGCCAGTTGTGGCTGCTAAATTGCTTGGTAAAC
CAGTCTTTATTCATGAATCAGATCGGTCAATGGGACTAGCAAACAAGATTGCCTACAAATTTGCAACTACC
ATGTATACCACTTTTGAGCAGGAAGACCAGTTGTCTAAAGTTAAACACCTTGGAGCGGTGACAAAGGTTTT
CAAAGATGCCAACCAAATGCCTGAATCAACTCAGTTAGAGGCGGTGAAAGAGTATTTTAGTAGAGACCTAA
AAACCCTCTTGTTTATTGGTGGTTCGGCAGGGGCGCATGTGTTTAATCAGTTTATTAGTGATCATCCAGAA
TTGAAGCAACGTTATAATATCATCAATATTACAGGAGACCCTCACCTTAATGAATTGAGTTCTCATCTGTA
TCGAGTAGATTATGTTACCGATCTCTACCAACCTTTGATGGCGATGGCTGACCTTGTAGTGACAAGAGGGG
GCTCTAATACACTTTTTGAGCTACTGGCAATGGCTAAGCTACACCTCATCGTTCCTCTTGGTAAAGAAGCT
AGCCGTGGCGATCAGTTAGAAAATGCCACTTATTTTGAGAAGAGGGGCTACGCTAAACAATTACAGGAACC
TGATTTAACTTTGCATAATTTTGATCAGGCAATGGCTGATTTGTTTGAACATCAGGCTGATTATGAGGCTA
CTATGTTGGCAACTAAGGAGATTCAGTCACCGGACTTCTTTTATGACCTTTTGAGAGCTGATATTAGCTCC
GCGATTAAGGAGAAGTAA

SEQ ID NO: 94 amino acid sequence comprising GAS 529

MCGIVGVVGNRNATDILMQGLEKLEYRGYDSAGIFVANANQTNLIKSVGRIADLRAKIGIDVAGSTGIGHT
RWATHGQSTEDNAHPHTSQTGRFVLVHNGVIENYLHIKTEFLAGHDFKGQTDTEIAVHLIGKFVEEDKLSV
LEAFKKSLSIIEGSYAFALMDSQATDTIYVAKNKSPLLIGLGEGYNMVCSDAMAMIRETSEFMEIHDKELV
ILTKDKVTVTDYDGKELIRDSYTAELDLSDIGKGTYPFYMLKEIDEQPTVMRQLISTYADETGNVQVDPAI
ITSIQEADRLYILAAGTSYHAGFATKNMLEQLTDTPVELGVASEWGYHMPLLSKKPMFILLSQSGETADSR
QVLVKANAMGIPSLTVTNVPGSTLSREATYTMLIHAGPEIAVASTKAYTAQIAALAFLAKAVGEANGKQEA
LDFNLVHELSLVAQSIEATLSEKDLVAEKVQALLATTRNAFYIGRGNDYYVAMEAALKLKEISYIQCEGFA
AGELKHGTISLIEEDTPVIALISSSQLVASHTRGNIQEVAARGAHVLTVVEEGLDREGDDIIVNKVHPFLA
PIAMVIPTQLIAYYASLQRGLDVDKPRNLAKAVTVE

SEQ ID NO: 95 polynucleotide sequence encoding GAS 529

ATGTGTGGAATTGTTGGAGTTGTTGGAAATCGCAATGCAACGGATATTTTAATGCAAGGCCTTGAAAAGCT
TGAATACCGGGGTTATGATTCAGCAGGAATTTTTGTGGCTAATGCCAATCAAACAAACTTGATTAAATCAG

SEQUENCE LISTING

TGGGGCGGATTGCTGATTTGCGTGCCAAGATTGGCATTGATGTTGCTGGTTCAACAGGGATTGGTCACACC
CGTTGGGCAACGCATGGCCAATCAACAGAGGATAATGCCCATCCTCACACGTCACAAACTGGACGTTTTGT
ACTTGTTCATAATGGTGTGATTGAAAATTACCTTCACATTAAAACAGAGTTCCTAGCTGGACATGATTTTA
AGGGGCAGACAGATACTGAGATTGCAGTACACTTGATTGGAAAATTTGTGGAAGAAGACAAGTTGTCAGTA
CTGGAAGCTTTTAAAAAATCTTTAAGCATTATTGAAGGTTCCTACGCCTTTGCATTAATGGATAGCCAAGC
AACTGATACTATTTATGTGGCTAAAAACAAGTCTCCATTGTTGATTGGACTTGGTGAAGGTTACAACATGG
TTTGTTCAGATGCCATGGCCATGATTCGTGAAACCAGTGAATTTATGGAAATTCATGATAAGGAGCTAGTT
ATTTTAACCAAAGATAAGGTAACTGTTACAGACTACGATGGTAAAGAGCTGATACGAGATTCCTACACTGC
TGAATTAGACTTATCTGATATTGGCAAAGGGACTTATCCTTTCTATATGCTGAAAGAAATTGATGAGCAAC
CAACCGTAATGCGTCAATTAATTTCAACTTATGCAGATGAAACTGGTAACGTACAGGTTGATCCGGCTATC
ATTACCTCTATCCAAGAGGCTGACCGTCTTTATATTTTAGCGGCAGGGACTTCCTACCATGCTGGTTTTGC
AACAAAAAATATGCTTGAGCAATTGACAGATACACCAGTTGAGTTGGGCGTGGCTTCTGAGTGGGGTTACC
ACATGCCTCTGCTTAGCAAGAAACCAATGTTTATTCTACTAAGCCAATCAGGAGAAACCGCAGATAGTCGT
CAAGTTTTAGTAAAGGCAAATGCTATGGGCATTCCGAGTTTGACAGTAACTAACGTTCCAGGATCAACCTT
ATCACGTGAAGCAACATACACCATGTTGATTCATGCTGGACCTGAAATTGCTGTTGCGTCTACAAAAGCTT
ACACTGCACAAATTGCTGCCCTTGCCTTTTTGGCTAAGGCAGTTGGTGAGGCAAATGGTAAGCAAGAAGCT
CTTGACTTTAACTTGGTACATGAGTTGTCATTGGTTGCCCAATCTATTGAGGCGACTTTGTCTGAAAAAGA
TCTCGTGGCAGAAAAGGTTCAAGCTTTGCTAGCTACTACTCGTAATGCTTTTTACATCGGGCGTGGCAATG
ATTATTACGTTGCGATGGAAGCTGCTTTGAAATTAAAAGAGATTTCTTATATTCAATGCGAAGGCTTTGCG
GCTGGTGAATTGAAACATGGAACCATTTCATTAATTGAGGAGGACACGCCAGTAATCGCTTTAATATCGTC
TAGTCAGTTGGTTGCCTCTCATACGCGTGGTAATATTCAAGAAGTTGCTGCCCGTGGGGCTCATGTTTTAA
CAGTTGTGGAAGAAGGGCTTGACCGTGAGGGAGATGACATTATTGTCAATAAGGTTCATCCTTTCCTAGCC
CCGATTGCTATGGTCATTCCAACTCAACTGATTGCTTACTACGCTTCATTACAACGTGGACTTGATGTTGA
TAAGCCACGTAATTTGGCTAAAGCTGTAACAGTAGAATAA

SEQ ID NO: 96 amino acid sequence comprising GAS 45

VTFMKKSKWLAAVSVAILSVSALAACGNKNASGGSEATKTYKYVFVNDPKSLDYILTNGGGTTDVITQMVD
GLLENDEYGNLVPSLAKDWKVSKDGLTYTYTLRDGVSWYTADGEEYAPVTAEDFVTGLKHAVDDKSDALYV
VEDSIKNLKAYQNGEVDFKEVGVKALDDKTVQYTLNKPESYWNSKTTYSVLFPVNAKFLKSKGKDFGTTDP
SSILVNGAYFLSAFTSKSSMEFHKNENYWDAKNVGIESVKLTYSDGSDPGSFYKNFDKGEFSVARLYPNDP
TYKSAKKNYADNITYGMLTGDIRHLTWNLNRTSFKNTKKDPAQQDAGKKALNNKDFRQAIQFAFDRASFQA
QTAGQDAKTKALRNMLVPPTFVTIGESDFGSEVEKEMAKLGDEWKDVNLADAQDGFYNPEKAKAEFAKAKE
ALTAEGVTFPVQLDYPVDQANAATVQEAQSFKQSVEASLGKENVIVNVLETETSTHEAQGFYAETPEQQDY
DIISSWWGPDYQDPRTYLDIMSPVGGGSVIQKLGIKAGQNKDVVAAAGLDTYQTLLDEAAAITDDNDARYK
AYAKAQAYLTDNAVDIPVVALGGTPRVTKAVPFSGGFSWAGSKGPLAYKGMKLQDKPVTVKQYEKAKEKWM
KAKAKSNAKYAEKLADHVEK

SEQ ID NO: 97 polynucleotide sequence encoding GAS 45

GTGACTTTTATGAAGAAAAGTAAATGGTTGGCAGCTGTAAGTGTTGCGATCTTGTCAGTATCCGCTTTGGC
AGCTTGTGGTAATAAAAATGCTTCAGGTGGCTCAGAAGCTACAAAAACCTACAAGTACGTTTTTGTTAACG
ATCCAAAATCATTGGATTATATTTTGACTAATGGCGGTGGAACGACTGATGTGATAACACAAATGGTTGAT
GGTCTTTTGGAAAACGATGAGTATGGTAATTTAGTACCATCACTTGCTAAAGATTGGAAGGTTTCAAAAGA
CGGTCTGACTTATACTTATACTCTTCGCGATGGTGTCTCTTGGTATACGGCTGATGGTGAAGAATATGCCC
CAGTAACAGCAGAAGATTTTGTGACTGGTTTGAAGCACGCGGTTGACGATAAATCAGATGCTCTTTACGTT
GTTGAAGATTCAATAAAAAACTTAAAGGCTTACCAAAATGGTGAAGTAGATTTTAAAGAAGTTGGTGTCAA
AGCCCTTGACGATAAAACTGTTCAGTATACTTTGAACAAGCCTGAAAGCTACTGGAATTCAAAAACAACTT
ATAGTGTGCTTTTCCCAGTTAATGCGAAATTTTTGAAGTCAAAAGGTAAAGATTTTGGTACAACCGATCCA
TCATCAATCCTTGTTAATGGTGCTTACTTCTTGAGCGCCTTCACCTCAAAATCATCTATGGAATTCCATAA
AAATGAAAACTACTGGGATGCTAAGAATGTTGGGATAGAATCTGTTAAATTGACTTACTCAGATGGTTCAG
ACCCAGGTTCGTTCTACAAGAACTTTGACAAGGGTGAGTTCAGCGTTGCACGACTTTACCCAAATGACCCT
ACCTACAAATCAGCTAAGAAAAACTATGCTGATAACATTACTTACGGAATGTTGACTGGAGATATCCGTCA
TTTAACATGGAATTTGAACCGTACTTCTTTCAAAAACACTAAGAAAGACCCTGCACAACAAGATGCCGGTA
AGAAAGCTCTTAACAACAAGGATTTTCGTCAAGCTATTCAGTTTGCTTTTGACCGAGCGTCATTCCAAGCA
CAAACTGCAGGTCAAGATGCCAAAACAAAAGCCTTACGTAACATGCTTGTCCCACCAACATTTGTGACCAT
TGGAGAAAGTGATTTTGGTTCAGAAGTTGAAAAGGAAATGGCAAAACTTGGTGATGAATGGAAAGACGTTA
ACTTAGCTGATGCTCAAGATGGTTTCTATAATCCTGAAAAAGCAAAAGCTGAGTTTGCAAAAGCCAAAGAA

SEQUENCE LISTING

GCTTTAACAGCTGAAGGTGTAACCTTCCCAGTTCAATTAGATTACCCTGTTGACCAAGCAAACGCAGCAAC
TGTTCAGGAAGCCCAGTCTTTCAAACAATCTGTTGAAGCATCTCTTGGTAAAGAGAATGTCATTGTCAATG
TTCTTGAAACAGAAACATCAACTCACGAAGCCCAAGGCTTCTATGCTGAGACCCCAGAACAACAAGACTAC
GATATCATTTCATCATGGTGGGGACCAGACTATCAAGATCCACGGACCTACCTTGACATCATGAGTCCAGT
AGGTGGTGGATCTGTTATCCAAAAACTTGGAATCAAAGCAGGTCAAAATAAGGATGTTGTGGCAGCTGCAG
GCCTTGATACCTACCAAACTCTTCTTGATGAAGCAGCAGCAATTACAGACGACAACGATGCGCGCTATAAA
GCTTACGCAAAAGCACAAGCCTACCTTACAGATAATGCCGTAGATATTCCAGTTGTGGCATTGGGTGGCAC
TCCACGAGTTACTAAAGCCGTTCCATTTAGCGGGGGCTTCTCTTGGGCAGGGTCTAAAGGTCCTCTAGCAT
ATAAAGGAATGAAACTTCAAGACAAACCTGTCACAGTAAAACAATACGAAAAAGCAAAAGAAAAATGGATG
AAAGCAAAGGCTAAGTCAAATGCAAAATATGCTGAGAAGTTAGCTGATCACGTTGAAAAA

SEQ ID NO: 98 amino acid sequence comprising an N-terminal leader sequence of GAS 45
VTFMKKSKWLAAVSVAILSVSALAA

SEQ ID NO: 99 amino acid sequence comprising a fragment of GAS 45 where the N-terminal leader sequence is removed
CGNKNASGGSEATKTYKYVFVNDPKSLDYILTNGGGTTDVITQMVDGLLENDEYGNLVPSLAKDWKVSKDG
LTYTYTLRDGVSWYTADGEEYAPVTAEDFVTGLKHAVDDKSDALYVVEDSIKNLKAYQNGEVDFKEVGVKA
LDDKTVQYTLNKPESYWNSKTTYSVLFPVNAKFLKSKGKDFGTTDPSSILVNGAYFLSAFTSKSSMEFHKN
ENYWDAKNVGIESVKLTYSDGSDPGSFYKNFDKGEFSVARLYPNDPTYKSAKKNYADNITYGMLTGDIRHL
TWNLNRTSFKNTKKDPAQQDAGKKALNNKDFRQAIQFAFDRASFQAQTAGQDAKTKALRNMLVPPTFVTIG
ESDFGSEVEKEMAKLGDEWKDVNLADAQDGFYNPEKAKAEFAKAKEALTAEGVTFPVQLDYPVDQANAATV
QEAQSFKQSVEASLGKENVIVNVLETETSTHEAQGFYAETPEQQDYDIISSWWGPDYQDPRTYLDIMSPVG
GGSVIQKLGIKAGQNKDVVAAAGLDTYQTLLDEAAAITDDNDARYKAYAKAQAYLTDNAVDIPVVALGGTP
RVTKAVPFSGGFSWAGSKGPLAYKGMKLQDKPVTVKQYEKAKEKWMKAKAKSNAKYAEKLADHVEK

SEQ ID NO: 100 amino acid sequence comprising GAS 95
MKIGKKIVLMFTAIVLTTVLALGVYLTSAYTFSTGELSKTFKDFSTSSNKSDAIKQTRAFSILLMGVDTGS
SERASKWEGNSDSMILVTVNPKTKKTTMTSLERDTLTTLSGPKNNEMNGVEAKLNAAYAAGGAQMAIMTVQ
DLLNITIDNYVQINMQGLIDLVNAVGGITVTNEFDFPISIAENEPEYQATVAPGTHKINGEQALVYARMRY
DDPEGDYGRQKRQREVIQKVLKKILALDSISSYRKILSAVSSNMQTNIEISSRTIPSLLGYRDALRTIKTY
QLKGEDATLSDGGSYQIVTSNHLLEIQNRIRTELGLHKVNQLKTNATVYENLYGSTKSQTVNNNYDSSGQA
PSYSDSHSSYANYSSGVDTGQSASTDQDSTASSHRPATPSSSSDALAADESSSSGSGSLVPPANINPQT

SEQ ID NO: 101 polynucleotide sequence encoding GAS 95
ATGAAAATTGGAAAAAAAATAGTTTTAATGTTCACAGCTATTGTGTTAACAACTGTCTTGGCATTAGGTGT
CTATCTAACTAGTGCTTATACCTTCTCAACAGGAGAATTATCAAAGACCTTTAAAGATTTTTCGACATCTT
CAAACAAAAGTGATGCCATTAAACAAACAAGAGCTTTTTCTATCTTGTTGATGGGTGTTGATACAGGCTCT
TCAGAGCGTGCCTCCAAGTGGGAAGGAAACAGTGATTCGATGATTTTGGTTACGGTTAATCCAAAGACCAA
GAAAACAACTATGACTAGTTTAGAACGAGATACCTTAACCACGTTATCTGGACCCAAAAATAATGAAATGA
ATGGTGTTGAAGCTAAGCTTAACGCTGCTTATGCAGCAGGTGGCGCTCAGATGGCTATTATGACCGTGCAA
GATCTTTTGAATATCACCATTGATAACTATGTTCAAATTAATATGCAAGGCCTTATTGATCTTGTGAATGC
AGTTGGAGGGATTACAGTTACAAATGAGTTTGATTTTCCTATCTCGATTGCTGAAAACGAACCTGAATATC
AAGCTACTGTTGCGCCTGGAACACACAAAATTAACGGTGAACAAGCTTTGGTTTATGCTCGTATGCGTTAT
GATGATCCTGAGGGAGATTATGGTCGACAAAAGCGTCAACGTGAAGTCATTCAAAAGGTATTGAAAAAAAT
CCTTGCTCTTGATAGCATTAGCTCTTATCGGAAGATTTTATCTGCTGTAAGTAGTAATATGCAAACGAATA
TCGAAATCTCTTCTCGCACTATCCCTAGTCTATTAGGTTATCGTGACGCACTTAGAACTATTAAGACTTAT
CAACTAAAAGGAGAAGATGCCACTTTATCAGATGGTGGATCATACCAAATTGTTACCTCTAATCATTTGTT
AGAAATCCAAAATCGTATCCGAACAGAATTAGGACTTCATAAGGTTAATCAATTAAAAACAAATGCTACTG
TTTATGAAAATTTGTATGGGTCAACTAAGTCTCAGACAGTAAACAACAACTATGACTCTTCAGGCCAGGCT
CCATCTTATTCTGATAGTCATAGCTCTTACGCTAATTATTCAAGTGGAGTAGATACCGGCCAGAGTGCTAG
TACAGACCAGGACTCTACTGCTTCAAGCCATAGGCCAGCTACGCCGTCTTCTTCATCAGATGCTTTAGCAG
CTGATGAGTCTAGCTCATCAGGGTCTGGATCATTAGTTCCTCCTGCTAATATCAACCCTCAGACCTAA

SEQ ID NO: 102 amino acid sequence comprising N-terminal leader sequence of GAS 95
MKIGKKIVLMFTAIVLTTVLALGVYLTSAYTFS

SEQUENCE LISTING

SEQ ID NO: 103 amino acid sequence comprising a fragment of GAS 95 where the N-terminal leader sequence is removed.

TGELSKTFKDFSTSSNKSDAIKQTRAFSILLMGVDTGSSERASKWEGNSDSMILVTVNPKTKKTTMTSLER
DTLTTLSGPKNNEMNGVEAKLNAAYAAGGAQMAIMTVQDLLNITIDNYVQINMQGLIDLVNAVGGITVTNE
FDFPISIAENEPEYQATVAPGTHKINGEQALVYARMRYDDPEGDYGRQKRQREVIQKVLKKILALDSISSY
RKILSAVSSNMQTNIEISSRTIPSLLGYRDALRTIKTYQLKGEDATLSDGGSYQIVTSNHLLEIQNRIRTE
LGLHKVNQLKTNATVYENLYGSTKSQTVNNNYDSSGQAPSYSDSHSSYANYSSGVDTGQSASTDQDSTASS
HRPATPSSSSDALAADESSSSGSGSLVPPANINPQT

SEQ ID NO: 104 amino acid sequence comprising GAS 193

MKKRKLLAVTLLSTILLNSAVPLVVADTSLRNSTSSTDQPTTADTDTDDESETPKKDKKSKETASQHDTQK
DHKPSHTHPTPPSNDTKQTDQASSEATDKPNKDKNDTKQPDSSDQSTPSPKDQSSQKESQNKDGRPTPSPD
QQKDQTPDKTPEKSADKTPEKGPEKATDKTPEPNRDAPKPIQPPLAAAPVFIPWRESDKDLSKLKPSSRSS
AAYVRHWTGDSAYTHNLLSRRYGITAEQLDGFLNSLGIHYDKERLNGKRLLEWEKLTGLDVRAIVAIAMAE
SSLGTQGVAKEKGANMFGYGAFDFNPNNAKKYSDEVAIRHMVEDTIIANKNQTFERQDLKAKKWSLGQLDT
LIDGGVYFTDTSGSGQRRADIMTKLDQWIDDHGSTPEIPEHLKITSGTQFSEVPVGYKRSQPQNVLTYKSE
TYSFGQCTWYAYNRVKELGYQVDRYMGNGGDWQRKPGFVTTHKPKVGYVVSFAPGQAGADATYGHVAVVEQ
IKEDGSILISESNVMGLGTISYRTFTAEQASLLTYVVGDKLPRP

SEQ ID NO: 105 polynucleotide sequence encoding GAS 193

ATGAAGAAAAGGAAATTGTTAGCAGTAACACTATTAAGTACCATACTCTTAAACAGTGCAGTGCCATTAGT
TGTTGCTGATACCTCCTTGCGTAATAGCACATCATCCACTGATCAGCCTACTACAGCAGATACTGATACGG
ATGACGAGAGTGAAACACCAAAAAAAGACAAAAAAAGCAAGGAAACAGCGTCGCAGCACGACACCCAAAAA
GACCATAAGCCATCACACACTCACCCAACCCCCCCTTCAAATGATACTAAGCAGACCGATCAGGCATCATC
TGAAGCTACTGACAAACCAAATAAAGACAAAAACGACACCAAGCAACCAGACAGCAGTGATCAATCCACCC
CATCTCCCAAAGACCAGTCGTCTCAAAAAGAGTCACAAAACAAAGACGGCCGACCTACCCCATCACCTGAT
CAGCAAAAAGATCAGACACCTGATAAAACACCAGAAAAATCAGCTGATAAAACCCCTGAAAAAGGACCAGA
AAAAGCAACTGATAAAACACCAGAGCCAAATCGTGACGCTCCAAAACCCATCCAACCTCCTTTAGCAGCTG
CTCCTGTCTTTATACCTTGGAGAGAAAGTGACAAAGACCTGAGCAAGCTAAAACCAAGCAGTCGCTCATCA
GCGGCTTACGTGAGACACTGGACAGGTGACTCTGCCTACACTCACAACCTGTTGTCACGCCGTTATGGGAT
TACTGCTGAACAGCTAGATGGTTTTTTGAACAGTCTAGGTATTCACTATGATAAAGAACGCTTAAACGGAA
AGCGTTTATTAGAATGGGAAAAACTAACAGGACTAGACGTTCGAGCTATCGTAGCTATTGCAATGGCAGAA
AGCTCACTAGGTACTCAGGGAGTTGCTAAAGAAAAAGGAGCCAATATGTTTGGTTATGGCGCCTTTGACTT
CAACCCAAACAATGCCAAAAAATACAGCGATGAGGTTGCTATTCGTCACATGGTAGAAGACACCATCATTG
CCAACAAAAACCAAACCTTTGAAAGACAAGACCTCAAAGCAAAAAAATGGTCACTAGGCCAGTTGGATACC
TTGATTGATGGTGGGGTTTACTTTACAGATACAAGTGGCAGTGGGCAAAGACGAGCAGATATCATGACCAA
ACTAGACCAATGGATAGATGATCATGGAAGCACACCTGAGATTCCAGAACATCTCAAGATAACTTCCGGGA
CACAATTTAGCGAAGTGCCCGTAGGTTATAAAAGAAGTCAGCCACAAAACGTTTTGACCTACAAGTCAGAG
ACCTACAGCTTTGGCCAATGCACTTGGTACGCCTATAATCGTGTCAAAGAGCTAGGTTATCAAGTCGACAG
GTACATGGGTAACGGTGGCGACTGGCAGCGCAAGCCAGGTTTTGTGACCACCCATAAACCTAAAGTGGGCT
ATGTCGTCTCATTTGCACCAGGCCAAGCAGGAGCAGATGCAACCTATGGTCACGTTGCTGTTGTAGAGCAA
ATCAAAGAAGATGGTTCTATCTTAATTTCAGAGTCAAATGTTATGGGACTAGGCACCATTTCCTATCGGAC
GTTCACAGCTGAGCAGGCTAGTTTGTTGACCTATGTCGTAGGGGACAAACTCCCAAGACCATAA

SEQ ID NO: 106 amino acid sequence comprising GAS 137

MSDKHINLVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPPALVPKFLELIEQTNENRRVALVVDMRSRLFFK
EINSTLDSIESNPSIDFRILFLDATDGELVSRYKETRRSHPLAADGRVLDGIRLERELLSPLKSMSQHVVD
TTKLTPRQLRKTISDQFSEGSNQASFRIEVMSFGFKYGLPLDADLVFDVRFLPNPYYQVELREKTGLDEDV
FNYVMSHPESEVFYKHLLNLIVPILPAYQKEGKSVLTVAIGCTGGQHRSVAFAHCLAESLATDWSVNESHR
DQNRRKETVNRS

SEQ ID NO: 107 polynucleotide sequence encoding GAS 137

ATGTCAGACAAACACATTAATTTAGTTATTGTGACAGGAATGAGCGGCGCTGGAAAAACAGTTGCCATTCA
GTCTTTTGAGGATCTAGGCTACTTTACCATTGATAATATGCCCCCAGCCTTGGTTCCAAAATTTTTAGAAT
TAATTGAACAAACCAATGAAAATCGTAGGGTGGCTTTGGTTGTCGATATGAGAAGTCGTTTGTTTTTCAAG

SEQUENCE LISTING

GAAATTAATTCTACCTTAGATAGTATTGAAAGCAATCCTAGCATTGATTTTCGGATTCTTTTTTTGGATGC
AACGGATGGAGAATTGGTGTCACGCTATAAAGAAACCAGACGGAGCCACCCTTTGGCTGCGGACGGTCGTG
TGCTTGATGGTATTCGATTGGAAAGAGAACTCCTATCTCCTTTGAAAAGCATGAGCCAACATGTGGTGGAT
ACAACAAAATTGACCCCTAGACAATTGCGTAAAACCATTTCAGACCAGTTTTCTGAAGGGTCTAATCAAGC
CTCTTTCCGTATTGAAGTGATGAGCTTTGGGTTCAAATATGGTCTTCCTTTGGATGCGGATTTGGTTTTTG
ATGTGCGTTTTCTACCCAATCCTTATTATCAGGTAGAGCTTCGTGAAAAAACAGGACTAGATGAGGACGTT
TTTAATTATGTGATGTCTCACCCAGAATCAGAGGTGTTTTACAAGCATTTGTTAAACCTTATTGTCCCTAT
CTTACCGGCTTACCAAAAAGAAGGGAAGTCTGTCTTGACGGTGGCTATTGGCTGCACAGGAGGCCAACACC
GCAGCGTTGCCTTTGCCCATTGCTTGGCAGAAAGTCTGGCAACAGATTGGTCGGTTAATGAAAGCCATCGT
GATCAAAATCGTCGTAAGGAAACGGTGAATCGTTCATGA

SEQ ID NO: 108 amino acid sequence comprising GAS 84

MIIKKRTVAILAIASSFFLVACQATKSLKSGDAWGVYQKQKSITVGFDNTFVPMGYKDESGRCKGFDIDLA
KEVFHQYGLKVNFQAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAFTDSYMRNEQIIVVKKRSDIKTI
SDMKHKVLGAQSASSGYDSLLRTPKLLKDFIKNKDANQYETFTQAFIDLKSDRIDGILIDKVYANYYLAKE
GQLENYRMIPTTFENEAFSVGLRKEDKTLQAKINRAFRVLYQNGKFQAISEKWFGDDVATANIKS

SEQ ID NO: 109 polynucleotide sequence encoding GAS 84

ATGATTATAAAAAAAAGAACCGTAGCAATTTTAGCCATAGCTAGTAGCTTTTTCTTGGTAGCTTGTCAAGC
TACTAAAAGTCTTAAATCAGGAGATGCTTGGGGAGTTTACCAAAAGCAAAAAAGTATTACAGTTGGTTTTG
ACAATACGTTTGTTCCTATGGGCTATAAGGATGAAAGCGGCAGATGCAAAGGTTTTGATATTGATTTGGCT
AAAGAAGTTTTTCACCAATATGGACTCAAGGTTAACTTTCAAGCTATTAATTGGGACATGAAAGAAGCAGA
ACTAAACAATGGTAAAATTGATGTAATCTGGAATGGTTATTCAATAACTAAGGAGCGTCAGGATAAGGTTG
CCTTTACTGATTCTTACATGAGAAATGAACAAATTATTGTTGTCAAAAAAAGATCTGATATTAAAACAATA
TCAGATATGAAACATAAAGTGTTAGGAGCACAATCAGCTTCATCAGGTTATGACTCCTTGTTAAGAACTCC
TAAACTGCTGAAAGATTTTATTAAAAATAAAGACGCTAATCAATATGAAACCTTTACACAAGCTTTTATTG
ATTTAAAATCAGATCGTATCGATGGAATATTGATTGACAAAGTATATGCCAATTACTATTTAGCAAAAGAA
GGGCAATTAGAGAATTATCGGATGATCCCAACGACCTTTGAAAATGAAGCATTTTCGGTTGGACTTAGAAA
AGAAGACAAAACGTTGCAAGCAAAAATTAATCGTGCTTTCAGGGTGCTTTATCAAAATGGCAAATTTCAAG
CTATTTCTGAGAAATGGTTTGGAGATGATGTTGCCACTGCCAATATTAAATCTTAA

SEQ ID NO: 110 amino acid sequence comprising N-terminal leader sequence of GAS 84

MIIKKRTVAILAIASSFFLVA

SEQ ID NO: 111 amino acid sequence comprising a fragment of GAS 84 where the N-terminal leader sequence is removed

CQATKSLKSGDAWGVYQKQKSITVGFDNTFVPMGYKDESGRCKGFDIDLAKEVFHQYGLKVNFQAINWDMK
EAELNNGKIDVIWNGYSITKERQDKVAFTDSYMRNEQIIVVKKRSDIKTISDMKHKVLGAQSASSGYDSLL
RTPKLLKDFIKNKDANQYETFTQAFIDLKSDRIDGILIDKVYANYYLAKEGQLENYRMIPTTFENEAFSVG
LRKEDKTLQAKINRAFRVLYQNGKFQAISEKWFGDDVATANIKS

SEQ ID NO: 112 amino acid sequence comprising GAS 384

MKTLAFDTSNKTLSLAILDDETLLADMTLNIQKKHSVSLMPAIDFLMTCTDLKPQDLERIVVAKGPGSYTG
LRVAVATAKTLAYSLNIALVGISSLYALAASTCKQYPNTLVVPLIDARRQNAYVGYYRQGKSVMPQAHASL
EVIIEQLVEEGQLIFVGETAPFAEKIQKKLPQAILLPTLPSAYECGLLGQSLAPENVDAFVPQYLKRVEAE
ENWLKDNEIKDDSHYVKRI

SEQ ID NO: 113 polynucleotide sequence encoding GAS 384

ATGAAGACACTTGCATTTGATACCTCAAATAAAACCTTGTCCCTTGCTATACTTGATGATGAGACACTTCT
AGCAGATATGACCCTTAACATTCAGAAAAAACATAGTGTTAGCCTTATGCCTGCTATTGATTTTTTGATGA
CTTGTACTGATCTTAAACCTCAAGATTTAGAAAGAATAGTGGTTGCAAAAGGCCCTGGATCTTACACAGGT
TTACGAGTGGCAGTTGCTACTGCAAAAACGTTAGCGTACAGTTTAAATATTGCATTGGTCGGGATTTCGAG
TCTATATGCTTTGGCTGCGTCTACTTGTAAACAGTATCCAAATACTTTGGTGGTGCCATTGATTGATGCTA
GAAGGCAAAATGCGTATGTAGGTTATTATCGGCAAGGAAAATCAGTGATGCCACAAGCCCATGCTTCACTA
GAAGTTATTATAGAACAATTAGTAGAAGAAGGACAGCTGATTTTTGTTGGGGAGACTGCTCCTTTTGCTGA

SEQUENCE LISTING

GAAAATTCAAAAGAAACTACCTCAGGCGATACTACTTCCAACCCTTCCTTCTGCTTACGAATGTGGTCTTT
TGGGGCAAAGTTTGGCACCAGAAAATGTAGACGCCTTTGTCCCTCAATATCTCAAGAGAGTGGAAGCTGAA
GAAAACTGGCTCAAAGATAATGAGATAAAAGATGATAGTCACTACGTTAAGCGAATCTAA

SEQ ID NO: 114 amino acid sequence comprising GAS 202

MLKRLWLILGPLLIAFVLVVITIFSFPTQLDHSIAQEKANAVAITDSSFKNGLIKRQALSDETCRFVPFFG
SSEWSRMDSMHPSVLAERYKRSYRPFLIGKRGSASLSHYYGIQQITNEMQKKKAIFVVSPQWFTAQGINPS
AVQMYLSNTQVIEFLLKARTDKESQFAAKRLLELNPGVSKSNLLKKVSKGKSLSRLDRAILKCQHQVALRE
ESLFSFLGKSTNYEKRILPRVKGLPKVFSYKQLNALATKRGQLATTNNRFGIKNTFYRKRIAPKYNLYKNF
QVNYSYLASPEYNDFQLLLSEFAKRKTDVLFVITPVNKAWADYTGLNQDKYQAAVRKIKFQLKSQGFHRIA
DFSKDGGESYFMQDTIHLGWNGWLAFDKKVQPFLETKQPVPNYKMNPYFYSKIWANRKDLQ

SEQ ID NO: 115 polynucleotide sequence encoding GAS 202

ATGCTTAAGAGACTCTGGTTAATTCTAGGTCCTCTTCTTATTGCCTTTGTTTTAGTAGTGATTACTATTTT
TAGTTTTCCTACACAACTTGATCATTCCATAGCTCAGGAAAAAGCAAATGCCGTTGCGATCACAGATAGTT
CTTTTAAAAATGGTTTGATTAAAAGACAAGCTTTATCAGATGAGACTTGTCGTTTTGTGCCTTTTTTTGGT
TCTAGCGAATGGAGTCGAATGGATAGTATGCACCCTTCGGTGCTTGCAGAGCGCTACAAGCGGAGCTATAG
ACCATTTTTAATTGGTAAGAGAGGATCAGCATCTTTGTCGCATTATTATGGTATACAACAAATTACCAATG
AAATGCAAAAGAAAAAAGCCATCTTTGTAGTATCTCCTCAATGGTTTACTGCTCAAGGGATTAATCCTAGT
GCGGTTCAGATGTACTTGTCTAACACTCAAGTGATTGAATTTTTACTAAAAGCTAGAACTGATAAAGAATC
ACAGTTTGCAGCAAAGCGTTTGCTTGAGCTTAACCCTGGTGTGTCTAAATCAAACTTATTGAAAAAAGTAA
GTAAGGGTAAGTCTCTTAGTCGGTTAGACAGAGCTATTTTGAAATGTCAACATCAAGTAGCATTGAGAGAA
GAGTCCCTTTTTAGTTTTTTAGGCAAATCTACTAACTATGAAAAAAGAATTTTGCCTCGCGTTAAGGGATT
ACCTAAAGTATTTTCGTATAAACAATTGAATGCATTAGCAACTAAGAGAGGCCAATTAGCAACAACCAACA
ACCGTTTTGGGATTAAAAATACATTTTATCGTAAACGAATAGCACCTAAATACAATCTTTATAAGAATTTC
CAAGTTAATTATAGTTACCTGGCGTCACCAGAATACAATGATTTTCAGCTTTTATTATCAGAATTTGCTAA
ACGAAAAACAGATGTACTCTTTGTTATAACTCCTGTTAATAAAGCTTGGGCGGATTATACCGGCTTAAATC
AAGATAAGTATCAAGCGGCAGTTCGTAAAATAAAATTCCAGTTAAAGTCACAAGGATTTCATCGCATTGCT
GACTTCTCAAAAGATGGTGGTGAGTCCTACTTTATGCAAGATACCATCCATCTCGGTTGGAATGGCTGGTT
AGCTTTTGATAAGAAAGTGCAACCATTTCTAGAAACGAAGCAGCCAGTGCCCAACTATAAAATGAACCCTT
ATTTTTATAGTAAAATTTGGGCAAATAGGAAAGACTTGCAATAG

SEQ ID NO: 116 amino acid sequence comprising GAS 057

MEKKQRFSLRKYKSGTFSVLIGSVFLVMTTTVAADELSTMSEPTITNHAQQQAQHLTNTELSSAESKSQDT
SQITLKTNREKEQSQDLVSEPTTTELADTDAASMANTGSDATQKSASLPPVNTDVHDWVKTKGAWDKGYKG
QGKVVAVIDTGIDPAHQSMRISDVSTAKVKSKEDMLARQKAAGINYGSWINDKVVFAHNYVENSDNIKENQ
FEDFDEDWENFEFDAEAEPKAIKKHKIYRPQSTQAPKETVIKTEETDGSHDIDWTQTDDDTKYESHGMHVT
GIVAGNSKEAAATGERFLGIAPEAQVMFMRVFANDIMGSAESLFIKAIEDAVALGADVINLSLGTANGAQL
SGSKPLMEAIEKAKKAGVSVVVAAGNERVYGSDHDDPLATNPDYGLVGSPSTGRTPTSVAAINSKWVIQRL
MTVKELENRADLNHGKAIYSESVDFKDIKDSLGYDKSHQFAYVKESTDAGYNAQDVKGKIALIERDPNKTY
DEMIALAKKHGALGVLIFNNKPGQSNRSMRLTANGMGIPSAFISHEFGKAMSQLNGNGTGSLEFDSVVSKA
PSQKGNEMNHFSNWGLTSDGYLKPDITAPGGDIYSTYNDNHYGSQTGTSMASPQIAGASLLVKQYLEKTQP
NLPKEKIADIVKNLLMSNAQIHVNPETKTTTSPRQQGAGLLNIDGAVTSGLYVTGKDNYGSISLGNITDTM
TFDVTVHNLSNKDKTLRYDTELLTDHVDPQKGRFTLTSHSLKTYQGGEVTVPANGKVTVRVTMDVSQFTKE
LTKQMPNGYYLEGFVRFRDSQDDQLNRVNIPFVGFKGQFENLAVAEESIYRLKSQGKTGFYFDESGPKDDI
YVGKHFTGLVTLGSETNVSTKTISDNGLHTLGTFKNADGKFILEKNAQGNPVLAISPNGDNNQDFAAFKGV
FLRKYQGLKASVYHASDKEHKNPLWVSPESFKGDKNFNSDIRFAKSTTLLGTAFSGKSLTGAELPDGHYHY
VVSYYPDVVGAKRQEMTFDMILDRQKPVLSQATFDPETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYT
VTINDSYKYVSVEDNKTFVERQADGSFILPLDKAKLGDFYYMVEDFAGNVAIAKLGDHLPQTLGKTPIKLK
LTDGNYQTKETLKDNLEMTQSDTGLVTNQAQLAVVHRNQPQSQLTKMNQDFFISPNEDGNKDFVAFKGLKN
NVYNDLTVNVYAKDDHQKQTPIWSSQAGASVSAIESTAWYGITARGSKVMPGDYQYVVTYRDEHGKEHQKQ
YTISVNDKKPMITQGRFDTINGVDHFTPDKTKALDSSGIVREEVFYLAKKNGRKFDVTEGKDGITVSDNKV
YIPKNPDGSYTISKRDGVTLSDYYYLVEDRAGNVSFATLRDLKAVGKDKAVVNFGLDLPVPEDKQIVNFTY
LVRDADGKPIENLEYYNNSGNSLILPYGKYTVELLTYDTNAAKLESDKIVSFTLSADNNFQQVTFKITMLA
TSQITAHFDHLLPEGSRVSLKTAQDQLIPLEQSLYVPKAYGKTVQEGTYEVVVSLPKGYRIEGNTKVNTLP

SEQUENCE LISTING

NEVHELSLRLVKVGDASDSTGDHKVMSKNNSQALTASATPTKSTTSATAKALPSTGEKMGLKLRIVGLVLL
GLTCVFSRKKSTKD

SEQ ID NO: 117 polynucleotide sequence encoding GAS 057

GTGGAGAAAAAGCAACGTTTTTCCCTTAGAAAATACAAATCAGGAACGTTTTCGGTCTTAATAGGAAGCGT
TTTCTTGGTGATGACAACAACAGTAGCAGCAGATGAGCTAAGCACAATGAGCGAACCAACAATCACGAATC
ACGCTCAACAACAAGCGCAACATCTCACCAATACAGAGTTGAGCTCAGCTGAATCAAAATCTCAAGACACA
TCACAAATCACTCTCAAGACAAATCGTGAAAAAGAGCAATCACAAGATCTAGTCTCTGAGCCAACCACAAC
TGAGCTAGCTGACACAGATGCAGCATCAATGGCTAATACAGGTTCTGATGCGACTCAAAAAAGCGCTTCTT
TACCGCCAGTCAATACAGATGTTCACGATTGGGTAAAAACCAAAGGAGCTTGGGACAAGGGATACAAAGGA
CAAGGCAAGGTTGTCGCAGTTATTGACACAGGGATCGATCCGGCCCATCAAAGCATGCGCATCAGTGATGT
ATCAACTGCTAAAGTAAAATCAAAAGAAGACATGCTAGCACGCCAAAAAGCCGCCGGTATTAATTATGGGA
GTTGGATAAATGATAAAGTTGTTTTTGCACATAATTATGTGGAAAATAGCGATAATATCAAAGAAAATCAA
TTCGAGGATTTTGATGAGGACTGGGAAAACTTTGAGTTTGATGCAGAGGCAGAGCCAAAAGCCATCAAAAA
ACACAAGATCTATCGTCCCCAATCAACCCAGGCACCGAAAGAAACTGTTATCAAAACAGAAGAAACAGATG
GTTCACATGATATTGACTGGACACAAACAGACGATGACACCAAATACGAGTCACACGGTATGCATGTGACA
GGTATTGTAGCCGGTAATAGCAAAGAAGCCGCTGCTACTGGAGAACGCTTTTTAGGAATTGCACCAGAGGC
CCAAGTCATGTTCATGCGTGTTTTTGCCAACGACATCATGGGATCAGCTGAATCACTCTTTATCAAAGCTA
TCGAAGATGCCGTGGCTTTAGGAGCAGATGTGATCAACCTGAGTCTTGGAACCGCTAATGGGGCACAGCTT
AGTGGCAGCAAGCCTCTAATGGAAGCAATTGAAAAAGCTAAAAAAGCCGGTGTATCAGTTGTTGTAGCAGC
AGGAAATGAGCGCGTCTATGGATCTGACCATGATGATCCATTGGCGACAAATCCAGACTATGGTTTGGTCG
GTTCTCCCTCAACAGGTCGAACACCAACATCAGTGGCAGCTATAAACAGTAAGTGGGTGATTCAACGTCTA
ATGACGGTCAAAGAATTAGAAAACCGTGCCGATTTAAACCATGGTAAAGCCATCTATTCAGAGTCTGTCGA
CTTTAAAGACATAAAAGATAGCCTAGGTTATGATAAATCGCATCAATTTGCTTATGTCAAAGAGTCAACTG
ATGCGGGTTATAACGCACAAGACGTTAAAGGTAAAATTGCTTTAATTGAACGTGATCCCAATAAAACCTAT
GACGAAATGATTGCTTTGGCTAAGAAACATGGAGCTCTGGGAGTACTTATTTTTAATAACAAGCCTGGTCA
ATCAAACCGCTCAATGCGTCTAACAGCTAATGGGATGGGGATACCATCTGCTTTCATATCGCACGAATTTG
GTAAGGCCATGTCCCAATTAAATGGCAATGGTACAGGAAGTTTAGAGTTTGACAGTGTGGTCTCAAAAGCA
CCGAGTCAAAAAGGCAATGAAATGAATCATTTTTCAAATTGGGGCCTAACTTCTGATGGCTATTTAAAACC
TGACATTACTGCACCAGGTGGCGATATCTATTCTACCTATAACGATAACCACTATGGTAGCCAAACAGGAA
CAAGTATGGCCTCTCCTCAGATTGCTGGCGCCAGCCTTTTGGTCAAACAATACCTAGAAAAGACTCAGCCA
AACTTGCCAAAAGAAAAAATTGCTGATATCGTTAAGAACCTATTGATGAGCAATGCTCAAATTCATGTTAA
TCCAGAGACAAAAACGACCACCTCACCGCGTCAGCAAGGGGCAGGATTACTTAATATTGACGGAGCTGTCA
CTAGCGGCCTTTATGTGACAGGAAAAGACAACTATGGCAGTATATCATTAGGCAACATCACAGATACGATG
ACGTTTGATGTGACTGTTCACAACCTAAGCAATAAAGACAAAACATTACGTTATGACACAGAATTGCTAAC
AGATCATGTAGACCCACAAAAGGGCCGCTTCACTTTGACTTCTCACTCCTTAAAAACGTACCAAGGAGGAG
AAGTTACAGTCCCAGCCAATGGAAAAGTGACTGTAAGGGTTACCATGGATGTCTCACAGTTCACAAAAGAG
CTAACAAAACAGATGCCAAATGGTTACTATCTAGAAGGTTTTGTCCGCTTTAGAGATAGTCAAGATGACCA
ACTAAATAGAGTAAACATTCCTTTTGTTGGTTTTAAAGGGCAATTTGAAAACTTAGCAGTTGCAGAAGAGT
CCATTTACAGATTAAAATCTCAAGGCAAAACTGGTTTTTACTTTGATGAATCAGGTCCAAAAGACGATATC
TATGTCGGTAAACACTTTACAGGACTTGTCACTCTTGGTTCAGAGACCAATGTGTCAACCAAAACGATTTC
TGACAATGGTCTACACACACTTGGCACCTTTAAAAATGCAGATGGCAAATTTATCTTAGAAAAAAATGCCC
AAGGAAACCCTGTCTTAGCCATTTCTCCAAATGGTGACAACAACCAAGATTTTGCAGCCTTCAAAGGTGTT
TTCTTGAGAAAATATCAAGGCTTAAAAGCAAGTGTCTACCATGCTAGTGACAAGGAACACAAAAATCCACT
GTGGGTCAGCCCAGAAAGCTTTAAAGGAGATAAAAACTTTAATAGTGACATTAGATTTGCAAAATCAACGA
CCCTGTTAGGCACAGCATTTTCTGGAAAATCGTTAACAGGAGCTGAATTACCAGATGGGCATTATCATTAT
GTGGTGTCTTATTACCCAGATGTGGTCGGTGCCAAACGTCAAGAAATGACATTTGACATGATTTTAGACCG
ACAAAAACCGGTACTATCACAAGCAACATTTGATCCTGAAACAAACCGATTCAAACCAGAACCCCTAAAAG
ACCGTGGATTAGCTGGTGTTCGCAAAGACAGTGTCTTTTATCTAGAAAGAAAAGACAACAAGCCTTATACA
GTTACGATAAACGATAGCTACAAATATGTCTCAGTAGAAGACAATAAAACATTTGTGGAGCGACAAGCTGA
TGGCAGCTTTATCTTGCCGCTTGATAAAGCAAAATTAGGGGATTTCTATTACATGGTCGAGGATTTTGCAG
GGAACGTGGCCATCGCTAAGTTAGGAGATCACTTACCACAAACATTAGGTAAAACACCAATTAAACTTAAG
CTTACAGACGGTAATTATCAGACCAAAGAAACGCTTAAAGATAATCTTGAAATGACACAGTCTGACACAGG
TCTAGTCACAAATCAAGCCCAGCTAGCAGTGGTGCACCGCAATCAGCCGCAAAGCCAGCTAACAAAGATGA
ATCAGGATTTCTTTATCTCACCAAACGAAGATGGGAATAAAGACTTTGTGGCCTTTAAAGGCTTGAAAAAT
AACGTGTATAATGACTTAACGGTTAACGTATACGCTAAAGATGACCACCAAAAACAAACCCCTATCTGGTC

SEQUENCE LISTING

TAGTCAAGCAGGCGCTAGTGTATCCGCTATTGAAAGTACAGCCTGGTATGGCATAACAGCCCGAGGAAGCA
AGGTGATGCCAGGTGATTATCAGTATGTTGTGACCTATCGTGACGAACATGGTAAAGAACATCAAAAGCAG
TACACCATATCTGTGAATGACAAAAAACCAATGATCACTCAGGGACGTTTTGATACCATTAATGGCGTTGA
CCACTTTACTCCTGACAAGACAAAAGCCCTTGACTCATCAGGCATTGTCCGCGAAGAAGTCTTTTACTTGG
CCAAGAAAAATGGCCGTAAATTTGATGTGACAGAAGGTAAAGATGGTATCACAGTTAGTGACAATAAGGTG
TATATCCCTAAAAATCCAGATGGTTCTTACACCATTTCAAAAAGAGATGGTGTCACACTGTCAGATTATTA
CTACCTTGTCGAAGATAGAGCTGGTAATGTGTCTTTTGCTACCTTGCGTGACCTAAAAGCGGTCGGAAAAG
ACAAAGCAGTAGTCAATTTTGGATTAGACTTACCGGTCCCTGAAGACAAACAAATAGTGAACTTTACCTAC
CTTGTGCGGGATGCAGATGGTAAACCGATTGAAAACCTAGAGTATTATAATAACTCAGGTAACAGTCTTAT
CTTGCCATACGGCAAATACACGGTCGAATTGTTGACCTATGACACCAATGCAGCCAAACTAGAGTCAGATA
AAATCGTTTCCTTTACCTTGTCAGCTGATAACAACTTCCAACAAGTTACCTTTAAGATAACGATGTTAGCA
ACTTCTCAAATAACTGCCCACTTTGATCATCTTTTGCCAGAAGGCAGTCGCGTTAGCCTTAAAACAGCTCA
AGATCAGCTAATCCCGCTTGAACAGTCCTTGTATGTGCCTAAAGCTTATGGCAAAACCGTTCAAGAAGGCA
CTTACGAAGTTGTTGTCAGCCTGCCTAAAGGCTACCGTATCGAAGGCAACACAAAGGTGAATACCCTACCA
AATGAAGTGCACGAACTATCATTACGCCTTGTCAAAGTAGGAGATGCCTCAGATTCAACTGGTGATCATAA
GGTTATGTCAAAAAATAATTCACAGGCTTTGACAGCCTCTGCCACACCAACCAAGTCAACGACCTCAGCAA
CAGCAAAAGCCCTACCATCAACGGGTGAAAAAATGGGTCTCAAGTTGCGCATAGTAGGTCTTGTGTTACTC
GGACTTACTTGCGTCTTTAGCCGAAAAAAATCAACCAAAGATTGA

SEQ ID NO: 118 amino acid sequence comprising N-terminal leader sequence of GAS 57
MEKKQRFSLRKYKSGTFSVLIGSVFLVMTTTVA

SEQ ID NO: 119 amino acid sequence comprising a fragment of GAS 57 where the N-terminal leader sequence is removed
ADELSTMSEPTITNHAQQQAQHLTNTELSSAESKSQDTSQITLKTNREKEQSQDLVSEPTTTELADTDAAS
MANTGSDATQKSASLPPVNTDVHDWVKTKGAWDKGYKGQGKVVAVIDTGIDPAHQSMRISDVSTAKVKSKE
DMLARQKAAGINYGSWINDKVVFAHNYVENSDNIKENQFEDFDEDWENFEFDAEAEPKAIKKHKIYRPQST
QAPKETVIKTEETDGSHDIDWTQTDDDTKYESHGMHVTGIVAGNSKEAAATGERFLGIAPEAQVMFMRVFA
NDIMGSAESLFIKAIEDAVALGADVINLSLGTANGAQLSGSKPLMEAIEKAKKAGVSVVVAAGNERVYGSD
HDDPLATNPDYGLVGSPSTGRTPTSVAAINSKWVIQRLMTVKELENRADLNHGKAIYSESVDFKDIKDSLG
YDKSHQFAYVKESTDAGYNAQDVKGKIALIERDPNKTYDEMIALAKKHGALGVLIFNNKPGQSNRSMRLTA
NGMGIPSAFISHEFGKAMSQLNGNGTGSLEFDSVVSKAPSQKGNEMNHFSNWGLTSDGYLKPDITAPGGDI
YSTYNDNHYGSQTGTSMASPQIAGASLLVKQYLEKTQPNLPKEKIADIVKNLLMSNAQIHVNPETKTTTSP
RQQGAGLLNIDGAVTSGLYVTGKDNYGSISLGNITDTMTFDVTVHNLSNKDKTLRYDTELLTDHVDPQKGR
FTLTSHSLKTYQGGEVTVPANGKVTVRVTMDVSQFTKELTKQMPNGYYLEGFVRFRDSQDDQLNRVNIPFV
GFKGQFENLAVAEESIYRLKSQGKTGFYFDESGPKDDIYVGKHFTGLVTLGSETNVSTKTISDNGLHTLGT
FKNADGKFILEKNAQGNPVLAISPNGDNNQDFAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVSPESFKG
DKNFNSDIRFAKSTTLLGTAFSGKSLTGAELPDGHYHYVVSYYPDVVGAKRQEMTFDMILDRQKPVLSQAT
FDPETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYTVTINDSYKYVSVEDNKTFVERQADGSFILPLDK
AKLGDFYYMVEDFAGNVAIAKLGDHLPQTLGKTPIKLKLTDGNYQTKETLKDNLEMTQSDTGLVTNQAQLA
VVHRNQPQSQLTKMNQDFFISPNEDGNKDFVAFKGLKNNVYNDLTVNVYAKDDHQKQTPIWSSQAGASVSA
IESTAWYGITARGSKVMPGDYQYVVTYRDEHGKEHQKQYTISVNDKKPMITQGRFDTINGVDHFTPDKTKA
LDSSGIVREEVFYLAKKNGRKFDVTEGKDGITVSDNKVYIPKNPDGSYTISKRDGVTLSDYYYLVEDRAGN
VSFATLRDLKAVGKDKAVVNFGLDLPVPEDKQIVNFTYLVRDADGKPIENLEYYNNSGNSLILPYGKYTVE
LLTYDTNAAKLESDKIVSFTLSADNNFQQVTFKITMLATSQITAHFDHLLPEGSRVSLKTAQDQLIPLEQS
LYVPKAYGKTVQEGTYEVVVSLPKGYRIEGNTKVNTLPNEVHELSLRLVKVGDASDSTGDHKVMSKNNSQA
LTASATPTKSTTSATAKALPSTGEKMGLKLRIVGLVLLGLTCVFSRKKSTKD

SEQ ID NO: 120 amino acid sequence comprising C-terminal hydrophobic region
LPSTGEKMGLKLRIVGLVLLGLTCVFSRKKSTKD

SEQ ID NO: 121 amino acid sequence comprising a fragment of GAS 57 where the C-terminal hydrophobic region is removed
MEKKQRFSLRKYKSGTFSVLIGSVFLVMTTTVAADELSTMSEPTITNHAQQQAQHLTNTELSSAESKSQDT
SQITLKTNREKEQSQDLVSEPTTTELADTDAASMANTGSDATQKSASLPPVNTDVHDWVKTKGAWDKGYKG
QGKVVAVIDTGIDPAHQSMRISDVSTAKVKSKEDMLARQKAAGINYGSWINDKVVFAHNYVENSDNIKENQ

SEQUENCE LISTING

FEDFDEDWENFEFDAEAEPKAIKKHKIYRPQSTQAPKETVIKTEETDGSHDIDWTQTDDDTKYESHGMHVT
GIVAGNSKEAAATGERFLGIAPEAQVMFMRVFANDIMGSAESLFIKAIEDAVALGADVINLSLGTANGAQL
SGSKPLMEAIEKAKKAGVSVVVAAGNERVYGSDHDDPLATNPDYGLVGSPSTGRTPTSVAAINSKWVIQRL
MTVKELENRADLNHGKAIYSESVDFKDIKDSLGYDKSHQFAYVKESTDAGYNAQDVKGKIALIERDPNKTY
DEMIALAKKHGALGVLIFNNKPGQSNRSMRLTANGMGIPSAFISHEFGKAMSQLNGNGTGSLEFDSVVSKA
PSQKGNEMNHFSNWGLTSDGYLKPDITAPGGDIYSTYNDNHYGSQTGTSMASPQIAGASLLVKQYLEKTQP
NLPKEKIADIVKNLLMSNAQIHVNPETKTTTSPRQQGAGLLNIDGAVTSGLYVTGKDNYGSISLGNITDTM
TFDVTVHNLSNKDKTLRYDTELLTDHVDPQKGRFTLTSHSLKTYQGGEVTVPANGKVTVRVTMDVSQFTKE
LTKQMPNGYYLEGFVRFRDSQDDQLNRVNIPFVGFKGQFENLAVAEESIYRLKSQGKTGFYFDESGPKDDI
YVGKHFTGLVTLGSETNVSTKTISDNGLHTLGTFKNADGKFILEKNAQGNPVLAISPNGDNNQDFAAFKGV
FLRKYQGLKASVYHASDKEHKNPLWVSPESFKGDKNFNSDIRFAKSTTLLGTAFSGKSLTGAELPDGHYHY
VVSYYPDVVGAKRQEMTFDMILDRQKPVLSQATFDPETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYT
VTINDSYKYVSVEDNKTFVERQADGSFILPLDKAKLGDFYYMVEDFAGNVAIAKLGDHLPQTLGKTPIKLK
LTDGNYQTKETLKDNLEMTQSDTGLVTNQAQLAVVHRNQPQSQLTKMNQDFFISPNEDGNKDFVAFKGLKN
NVYNDLTVNVYAKDDHQKQTPIWSSQAGASVSAIESTAWYGITARGSKVMPGDYQYVVTYRDEHGKEHQKQ
YTISVNDKKPMITQGRFDTINGVDHFTPDKTKALDSSGIVREEVFYLAKKNGRKFDVTEGKDGITVSDNKV
YIPKNPDGSYTISKRDGVTLSDYYYLVEDRAGNVSFATLRDLKAVGKDKAVVNFGLDLPVPEDKQIVNFTY
LVRDADGKPIENLEYYNNSGNSLILPYGKYTVELLTYDTNAAKLESDKIVSFTLSADNNFQQVTFKITMLA
TSQITAHFDHLLPEGSRVSLKTAQDQLIPLEQSLYVPKAYGKTVQEGTYEVVVSLPKGYRIEGNTKVNTLP
NEVHELSLRLVKVGDASDSTGDHKVMSKNNSQALTASATPTKSTTSATAKA

SEQ ID NO: 122 amino acid sequence comprising a fragment of GAS 57 where both the N-terminal leader sequence and the C-terminal hydrophobic region are removed

ADELSTMSEPTITNHAQQQAQHLTNTELSSAESKSQDTSQITLKTNREKEQSQDLVSEPTTTELADTDAAS
MANTGSDATQKSASLPPVNTDVHDWVKTKGAWDKGYKGQGKVVAVIDTGIDPAHQSMRITSDVSTAKVKSKE
DMLARQKAAGINYGSWINDKVVFAHNYVENSDNIKENQFEDFDEDWENFEFDAEAEPKAIKKHKIYRPQST
QAPKETVIKTEETDGSHDIDWTQTDDDTKYESHGMHVTGIVAGNSKEAAATGERFLGIAPEAQVMFMRVFA
NDIMGSAESLFIKAIEDAVALGADVINLSLGTANGAQLSGSKPLMEAIEKAKKAGVSVVVAAGNERVYGSD
HDDPLATNPDYGLVGSPSTGRTPTSVAAINSKWVIQRLMTVKELENRADLNHGKAIYSESVDFKDIKDSLG
YDKSHQFAYVKESTDAGYNAQDVKGKIALIERDPNKTYDEMIALAKKHGALGVLIFNNKPGQSNRSMRLTA
NGMGIPSAFISHEFGKAMSQLNGNGTGSLEFDSVVSKAPSQKGNEMNHFSNWGLTSDGYLKPDITAPGGDI
YSTYNDNHYGSQTGTSMASPQIAGASLLVKQYLEKTQPNLPKEKIADIVKNLLMSNAQIHVNPETKTTTSP
RQQGAGLLNIDGAVTSGLYVTGKDNYGSISLGNITDTMTFDVTVHNLSNKDKTLRYDTELLTDHVDPQKGR
FTLTSHSLKTYQGGEVTVPANGKVTVRVTMDVSQFTKELTKQMPNGYYLEGFVRFRDSQDDQLNRVNIPFV
GFKGQFENLAVAEESIYRLKSQGKTGFYFDESGPKDDIYVGKHFTGLVTLGSETNVSTKTISDNGLHTLGT
FKNADGKFILEKNAQGNPVLAISPNGDNNQDFAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVSPESFKG
DKNFNSDIRFAKSTTLLGTAFSGKSLTGAELPDGHYHYVVSYYPDVVGAKRQEMTFDMILDRQKPVLSQAT
FDPETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYTVTINDSYKYVSVEDNKTFVERQADGSFILPLDK
AKLGDFYYMVEDFAGNVAIAKLGDHLPQTLGKTPIKLKLTDGNYQTKETLKDNLEMTQSDTGLVTNQAQLA
VVHRNQPQSQLTKMNQDFFISPNEDGNKDFVAFKGLKNNVYNDLTVNVYAKDDHQKQTPIWSSQAGASVSA
IESTAWYGITARGSKVMPGDYQYVVTYRDEHGKEHQKQYTISVNDKKPMITQGRFDTINGVDHFTPDKTKA
LDSSGIVREEVFYLAKKNGRKFDVTEGKDGITVSDNKVYIPKNPDGSYTISKRDGVTLSDYYYLVEDRAGN
VSFATLRDLKAVGKDKAVVNFGLDLPVPEDKQIVNFTYLVRDADGKPIENLEYYNNSGNSLILPYGKYTVE
LLTYDTNAAKLESDKIVSFTLSADNNFQQVTFKITMLATSQITAHFDHLLPEGSRVSLKTAQDQLIPLEQS
LYVPKAYGKTVQEGTYEVVVSLPKGYRIEGNTKVNTLPNEVHELSLRLVKVGDASDSTGDHKVMSKNNSQA
LTASATPTKSTTSATAKA

SEQ ID NO: 123 amino acid sequence of a GAS M protein

MAKNNTNRHYSLRKLKTGTASVAVALTVLGAGFANQTEVKANGDGNPREVIEDLAANNPAIQNIRLRYENK
DLKARLENAMEVAGRDFKRAEELEKAKQALEDQRKDLETKLKELQQDYDLAKESTSWDRQRLEKELEEKKE
ALELAIDQASRDYHRATALEKELEEKKKALELAIDQASQDYNRANVLEKELETITREQEINRNLLGNAKLE
LDQLSSEKEQLTIEKAKLEEEKQISDASRQSLRRDLDASREAKKQVEKDLANLTAELDKVKEDKQISDASR
QGLRRDLDASREAKKQVEKDLANLTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLA
ALEKLNKELEESKKLTEKEKAELQAKLEAEAKALKEQLAKQAEELAKLRAGKASDSQTPDTKPGNKAVPGK
GQAPQAGTKPNQNKAPMKETKRQLPSTGETANPFFTAAALTVMATAGVAAVVKRKEEN

SEQUENCE LISTING

SEQ ID NO: 124 amino acid sequence of GAS SfbI

MSFDGFFLHHLTNELKENLLYGRIQKVNQPFERELVLTIRNHRKNYKLLLLSAHPVFGRVQITQADFQNPQV
PNTFTMIMRKYLQGAVIEQLEQIDNDRIIEIKVSNKNEIGDAIQATLIIEIMGKHSNIILVDRAENKIIES
IKHVGFSQNSYRTILPGSTYIEPPKTAAVNPFTITDVPLFEILQTQELTVKSLQQHFQGLGRDTAKELAEL
LTTDKLKRFREFFARPTQANLTTASFAPVLFSDSHATFETLSDMLDHFYQDKAERDRINQQASDLIHRVQT
ELDKNRNKLSKQEAELLATENAELFRQKGELLTTYLSLVPNNQDSVILDNYYTGEKIEIALDKALTPNQNA
QRYFKKYQKLKEAVKHLSGLIADTKQSITYFESVDYNLSQASIDDIEDIREELYQAGFLKSRQRDKRHKRK
KPEQYLASDGTTILMVGRNNLQNEELTFKMAKKGELWFHAKDIPGSHVIIKDNLDPSDEVKTDAAELAAYY
SKARLSNLVQVDMIEAKKLHKPSGAKPGFVTYTGQKTLRVTPDQAKILSMKLS

SEQ ID NO: 125 amino acid sequence of a GAS Shp protein

MTKVVIKQLLQVIVVFMISLSTMTNLVYADKGQIYGCIIQRNYRHPISGQIEDSGGEHSFDIGQGMVEGTV
YSDAMLEVSDAGKIVLTFRMSLADYSGNYQFWIQPGGTGSFQAVDYNITQKGTDTNGTTLDIAISLPTVNS
IIRGSMFVEPMGREVVFYLSASELIQKYSGNMLAQLVTETDNSQNQEVKDSQKPVDTKLGESQDESHTGAM
ITQNKPKANSSNNKSLSDKKILPSKMGLTTSLELKKEDKFRSKKDLSIMIYYFPTFFLMLGGFAVWVWKKR
KKNDKTM

SEQ ID NO: 126 amino acids 10 to 30 of GAS protein SagA

FSIATGSGNSQGGSGSYTPGKC

SEQ ID NO: 127 polynucleotide sequence comprising fusion construct 117-40a-RR

ATGGCCTTTAACACAAGCCAGAGTGTCAGTGCACAAGTTTATAGCAATGAAGGGTATCACCAGCATTTGAC
TGATGAAAAATCACACCTGCAATATAGTAAAGACAACGCACAACTTCAATTGAGAAATATCCTTGACGGCT
ACCAAAATGACCTAGGGAGACACTACTCTAGCTATTATTACTACAACCTAAGAACCGTTATGGGACTATCA
AGTGAGCAAGACATTGAAAAACACTATGAAGAGCTTAAGAACAAGTTACATGATATGTACAATCATTATgc
tagcggtggcggatccATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAA
CGAAGGCGAGTAATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATT
GATGCAGTTGAAAAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAAC
TACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAA
TTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTA
ACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACA
AAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAA
ATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGAT
AATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAA
AAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTC
TTAAATCCTCAGCTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTAT
CCTCTTGAAGAACTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAA
AGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAG
CAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCA
GCTCACATGATTAATAGTGTAcGtcGtCAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGA
ATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGAC
AGCCAGGGGTATCAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCG
TCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAA
TGGTATTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACAT
ACGGCCATGCTATTAACTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCA
ACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACG
CTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTG
ATACTATTGCAGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCT
GATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGA
CAGCTTAAATCTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTA
AAGCAAAACAAGCACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCA
CTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCA
TTTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATA
CTAAGCAAGATTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCT
AAACAAAGCAGTCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAA
CGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTA

SEQUENCE LISTING

CGGGCGTAAAACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACG
AAACAACTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGG
CCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCAT
CTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAA
cGtgcggccgcactcgagCACCACCACCACCACCAC

SEQ ID NO: 128 amino acid sequence comprising fusion construct 117-40a-RR

M A F N T S Q S V S A Q V Y S N E G Y H Q H L T D E K S H L Q Y S K D N
A Q L Q L R N I L D G Y Q N D L G R H Y S S Y Y Y Y N L R T V **M** G L S S
E Q D I E K H Y E E L K N K L H D **M** Y N H Y A S G G G S **M** S V G V S H Q
V K A D D R A S G E T K A S N T H D D S L P K P E T I Q E A K A T I D A
V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H L K E Q Q D N E
Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E H Q R E L T A T
E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E T T R A Q D L V
E Q V K T S E Q N I A K L N A **M** I S N P D A I T K A A Q T A N D N T K A
L S S E L E K A K A D L E N Q K A K V K K Q L T E E L A A Q K A A L A E
K E A E L S R L K S S A P S T Q D S I V G N N T **M** K A P Q G Y P L E E L
K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K A S P G N Q L N
Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A Q F A A H **M** I N S
V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y K K T H G N T R P
S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A G A S G L I R N D
D N **M** Y E N I G A F N D V H T V N G I K R G I Y D S I K Y **M** L F T D H L
H G N T Y G H A I N F L R V D K H N P N A P V Y L G F S T S N V G S L N
E H F V **M** F P E S N I A N H Q R F N K T P I K A V G S T K D Y A Q R V G
T V S D T I A A I K G K V S S L E N R L S A I H Q E A D I **M** A A Q A K V
S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T K G S L R T E L L
A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L H Q T E A L A E Q
A A A R V T A L V A K K A H L Q Y L R D F K L N P N R L Q V I R E R I D
N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q S S L E A T I A T
T E H Q L T L L K T L A N E K E Y R H L D E D I A T V P D L Q V A P P L
T G V K P L S Y S K I D T T P L V Q E **M** V K E T K Q L L E A S A R L A A
E N T S L V A E A L V G Q T S E **M** V A S N A I V S K I T S S I T Q P S S
K T S Y G S G S S T T S N L I S D V D E S T Q R A A A L E H H H H H H

SEQ ID NO: 129 amino acid sequence comprising a linker in the 117-40a-RR construct
YASGGGS

SEQ ID NO: 130 polynucleotide sequence comprising 40a-RR-117 fusion construct
ATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATAC
TCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAA
CTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAAC
CACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCT
TGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAG
AGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCATTTCA
GCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAAATATTGCTAAGCTCAA
TGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAA
GCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAA
GAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAGCTCC
GTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAACTTA
AAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAA
ATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTT
TGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAATA
GTGTAcGtcGtCAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATTACTT
AGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCAGGGGTATCAGG
GCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATTCGAA
ATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACGTGGT

SEQUENCE LISTING

ATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTATTAA
CTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTAGGAT
CTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCT
ATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAGCGAT
CAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCTGATATTATGGCAGCCC
AAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAA
GTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAGCACA
ACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAAG
CCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCTAAGG
GACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTTGGC
TAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGTCTAG
AAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAACGAAAAGGAATATCGC
CACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAAACCGCT
ATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACTATTAGAAG
CTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGGCCAAACCTCTGAAATG
GTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGG
CTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAACgtgctagcggtggcg
gatccATGGCCTTTAACACAAGCCAGAGTGTCAGTGCACAAGTTTATAGCAATGAAGGGTATCACCAGCAT
TTGACTGATGAAAAATCACACCTGCAATATAGTAAAGACAACGCACAACTTCAATTGAGAAATATCCTTGA
CGGCTACCAAAATGACCTAGGGAGACACTACTCTAGCTATTATTACTACAACCTAAGAACCGTTATGGGAC
TATCAAGTGAGCAAGACATTGAAAAACACTATGAAGAGCTTAAGAACAAGTTACATGATATGTACAATCAT
TATgcggccgcactcgagCACCACCACCACCACCAC

SEQ ID NO: 131 amino acid sequence comprising the 40a-RR-117 fusion construct

M S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T I Q
E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H
L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E
H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E
T T R A Q D L V E Q V K T S E Q N I A K L N A **M** I S N P D A I T K A A Q
T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E L A
A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T **M** K A P
Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K
A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A Q
F A A H **M** I N S V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y K
K T H G N T R P S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A G
A S G L I R N D D N **M** Y E N I G A F N D V H T V N G I K R G I Y D S I K
Y **M** L F T D H L H G N T Y G H A I N F L R V D K H N P N A P V Y L G F S
T S N V G S L N E H F V **M** F P E S N I A N H Q R F N K T P I K A V G S T
K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E A D
I **M** A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T K
G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L H
Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N R L
Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q S
S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T V P
D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E **M** V K E T K Q L L
E A S A R L A A E N T S L V A E A L V G Q T S E **M** V A S N A I V S K I T
S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q R A S G G
G S **M** A F N T S Q S V S A Q V Y S N E G Y H Q H L T D E K S H L Q Y S K
D N A Q L Q L R N I L D G Y Q N D L G R H Y S S Y Y Y Y N L R T V **M** G L
S S E Q D I E K H Y E E L K N K L H D **M** Y N H Y A A A L E H H H H H H

SEQ ID NO: 132 polynucleotide sequence comprising fusion construct GAS 117 – 40a

ATGGCCTTTAACACAAGCCAGAGTGTCAGTGCACAAGTTTATAGCAATGAAGGGTATCACCAGCATTTGAC
TGATGAAAAATCACACCTGCAATATAGTAAAGACAACGCACAACTTCAATTGAGAAATATCCTTGACGGCT
ACCAAAATGACCTAGGGAGACACTACTCTAGCTATTATTACTACAACCTAAGAA**AC**[redacted]GTTATGGGACTATCA
AGTGAGCAAGACATTGAAAAACACTATGAAGAGCTTAAGAACAAGTTACATGATATGTACAATCATTATgc

SEQUENCE LISTING

tagcggtggcggatccATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAA
CGAAGGCGAGTAATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATT
GATGCAGTTGAAAAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAAC
TACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAA
TTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTA
ACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACA
AAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAA
ATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGAT
AATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAA
AAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTC
TTAAATCCTCAGCTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTAT
CCTCTTGAAGAACTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAA
AGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAG
CAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGTGGCTAGCGCAGTTTGCA
GCTCACATGATTAATAGTGTAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGA
ATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTACGGAC
AGCCAGGGGTATCAGGGCATTATGGTGTTGGGCCTCATGATAAAACTATTATTGAAGACTCTGCCGGAGCG
TCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAA
TGGTATTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACAT
ACGGCCATGCTATTAACTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTTCA
ACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACG
CTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTG
ATACTATTGCAGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCGGCTATTCATCAAGAAGCT
GATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGA
CAGCTTAAATCTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTA
AAGCAAAACAAGCACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCA
CTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCA
TTTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATA
CTAAGCAAGATTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCT
AAACAAAGCAGTCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAAACCTTAGCTAA
CGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTA
CGGGCGTAAAACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACG
AAACAACTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTTGG
CCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCAT
CTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAA
cGtgcggccgcactcgagCACCACCACCACCACCAC

SEQ ID NO: 133 amino acid sequence comprising fusion construct GAS 117-40a

MAFNTSQSVSAQVYSNEGYHQHLTDEKSHLQYSKDN
AQLQLRNILDGYQNDLGRHYSSYYYYNLRTVMGLSS
EQDIEKHYEELKNKLHDMYNHYASGGGSMSVGVSHQ
VKADDRASGETKASNTHDDSLPKPETIQEAKATIDA
VEKTLSQQKAELTELATALTKTTAEINHLKEQQDNE
QKALTSAQEIYTNTLASSEETLLAQGAEHQRELTAT
ETELHNAQADQHSKETALSEQKASISAETTRAQDLV
EQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTKA
LSSELEKAKADLENQKAKVKKQLTEELAAQKAALAE
KEAELSRLKSSAPSTQDSIVGNNTMKAPQGYPLEEL
KKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLN
QYQDIPADRNRFVDPDNLTPEVQNGLAQFAAHMINS
VRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRP
SFVYGQPGVSGHYGVGPHDKTIIEDSAGASGLIRND
DNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHL
HGNTYGHAINFLRVDKHNPNAPVYLGFSTSNVGSLN
EHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVG
TVSDTIAAIKGKVSSLENRLSAIHQEADIMAAQAKV

SEQUENCE LISTING

S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T K G S L R T E L L
A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L H Q T E A L A E Q
A A A R V T A L V A K K A H L Q Y L R D F K L N P N R L Q V I R E R I D
N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q S S L E A T I A T
T E H Q L T L L K T L A N E K E Y R H L D E D I A T V P D L Q V A P P L
T G V K P L S Y S K I D T T P L V Q E **M** V K E T K Q L L E A S A R L A A
E N T S L V A E A L V G Q T S E **M** V A S N A I V S K I T S S I T Q P S S
K T S Y G S G S S T T S N L I S D V D E S T Q R A A A L E H H H H H H

SEQ ID NO: 134 polynucleotide sequence comprising fusion construct GAS 117-40N

ATGGCCTTTAACACAAGCCAGAGTGTCAGTGCACAAGTTTATAGCAATGAAGGGTATCACCAGCATTTGAC
TGATGAAAAATCACACCTGCAATATAGTAAAGACAACGCACAACTTCAATTGAGAAATATCCTTGACGGCT
ACCAAAATGACCTAGGAGACACTACTCTAGCTATTATTACTACAACCTAAGAACCGTTATGGGACTATCA
AGTGAGCAAGACATTGAAAAACACTATGAAGAGCTTAAGAACAAGTTACATGATATGTACAATCATTATGC
tagcggtggcggatccATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAA
CGAAGGCGAGTAATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATT
GATGCAGTTGAAAAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAAC
TACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAA
TTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTA
ACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACA
AAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGTCTGAACAAA
ATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGAT
AATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAA
AAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTC
TTAAATCCTCAGCTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTAT
CCTCTTGAAGAACTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAA
AGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAgcggccgcac
tcgagCACCACCACCACCACCAC

SEQ ID NO: 135

M A F N T S Q S V S A Q V Y S N E G Y H Q H L T D E K S H L Q Y S K D N
A Q L Q L R N I L D G Y Q N D L G R H Y S S Y Y Y Y N L R T V **M** G L S S
E Q D I E K H Y E E L K N K L H D **M** Y N H Y A S G G G S **M** S V G V S H Q
V K A D D R A S G E T K A S N T H D D S L P K P E T I Q E A K A T I D A
V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H L K E Q Q D N E
Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E H Q R E L T A T
E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E T T R A Q D L V
E Q V K T S E Q N I A K L N A **M** I S N P D A I T K A A Q T A N D N T K A
L S S E L E K A K A D L E N Q K A K V K K Q L T E E L A A Q K A A L A E
K E A E L S R L K S S A P S T Q D S I V G N N T **M** K A P Q G Y P L E E L
K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K A S P G N Q L N
Q Y Q A A A L E H H H H H H

SEQ ID NO: 136

AGTTGGTA

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 April 2005 (14.04.2005)

PCT

(10) International Publication Number
WO 2005/032582 A3

(51) International Patent Classification[7]: A61K 39/09

(21) International Application Number:
PCT/US2004/024868

(22) International Filing Date: 30 July 2004 (30.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/491,822 31 July 2003 (31.07.2003) US
60/541,565 3 February 2004 (03.02.2004) US

(71) Applicant *(for all designated States except US)*: **CHIRON CORPORATION** [US/US]; 4560 Horton Street, Emeryville, CA 94608-2916 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **GRANDI, Guido** [IT/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **TELFORD, John** [IT/US]; C/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **BENSI, Giuliano** [IT/US]; C/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(74) Agents: **HALE, Rebecca, M.** et al.; Intellectual Property-R-338, Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
3 November 2005

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMMUNOGENIC COMPOSITIONS FOR STREPTOCOCCUS PYOGENES

(57) Abstract: The invention includes a GAS antigen, GAS 40, which is particularly suitable for use either alone or in combinations with additional GAS antigens, such as GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057.

WO 2005/032582 A3

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/024868

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 A61K39/09

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, Sequence Search, WPI Data, PAJ, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE Geneseq 'Online! 2 July 2002 (2002-07-02), "Streptococcus polypeptide SEQ ID NO 9188." XP002334315 retrieved from EBI accession no. GSN:ABP30006 Database accession no. ABP30006 the whole document	24-26
A	-& DATABASE EPO Proteins 'Online! 2 February 2004 (2004-02-02), "Sequence 7466 from Patent WO0234771." XP002334316 retrieved from EBI accession no. EPOP:CQ650509 Database accession no. CQ650509 the whole document	3,10-17
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 July 2005	23/08/2005

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Perez, C

4

Form PCT/ISA/210 (second sheet) (January 2004)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	-& WO 02/34771 A (CHIRON S.P.A; THE INSTITUTE FOR GENOMIC RESEARCH; TELFORD, JOHN; MASIG) 2 May 2002 (2002-05-02) claims 1-28 page 2, line 1 - page 10, line 16	1-17, 21-27
X	DATABASE Geneseq 'Online! 19 June 2003 (2003-06-19), "Protein encoded by Prokaryotic essential gene #31978." XP002334512 retrieved from EBI accession no. GSN:ABU46451 Database accession no. ABU46451 the whole document	24-26
X	-& WO 02/077183 A (ELITRA PHARMACEUTICALS, INC; WANG, LIANGSU; ZAMUDIO, CARLOS; MALONE, C) 3 October 2002 (2002-10-03) claim 25	24-26
X	DATABASE Geneseq 'Online! 2 December 2004 (2004-12-02), "S. pyogenes hyperimmune system reactive antigen Spy0269." XP002334513 retrieved from EBI accession no. GSN:ADR83896 Database accession no. ADR83896 the whole document	24-26
E	-& WO 2004/078907 A (INTERCELL AG; MEINKE, ANDREAS; NAGY, ESZTER; WINKLER, BIRGIT; GELBMANN) 16 September 2004 (2004-09-16) claim 11	24-26
A	FERRETTI J J ET AL: "Complete genome sequence of an M1 strain of Streptococcus pyogenes" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 98, no. 8, 10 April 2001 (2001-04-10), pages 4658-4663, XP002168716 ISSN: 0027-8424 the whole document	1-27

-/--

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/024868

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	OLIVE C ET AL: "Protection of mice from group A streptococcal infection by intranasal immunisation with a peptide vaccine that contains a conserved M protein B cell epitope and lacks a T cell autoepitope" VACCINE, BUTTERWORTH SCIENTIFIC. GUILDFORD, GB, vol. 20, no. 21-22, 21 June 2002 (2002-06-21), pages 2816-2825, XP004357806 ISSN: 0264-410X the whole document	1-27

4

Form PCT/ISA/210 (continuation of second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/024868

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0234771 A	02-05-2002	AU 1412702 A	06-05-2002
		CA 2425303 A1	02-05-2002
		EP 1328543 A2	23-07-2003
		WO 0234771 A2	02-05-2002
		MX PA03003690 A	05-05-2004
		ZA 200302739 A	17-11-2004
WO 02077183 A	03-10-2002	US 2002061569 A1	23-05-2002
		WO 02077183 A2	03-10-2002
		US 2004029129 A1	12-02-2004
WO 2004078907 A	16-09-2004	WO 2004078907 A2	16-09-2004

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

REVISED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 April 2005 (14.04.2005)

PCT

(10) International Publication Number
WO 2005/032582 A3

(51) International Patent Classification[7]: A61K 39/09

(21) International Application Number:
PCT/US2004/024868

(22) International Filing Date: 30 July 2004 (30.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/491,822 31 July 2003 (31.07.2003) US
60/541,565 3 February 2004 (03.02.2004) US

(71) Applicant *(for all designated States except US)*: **CHIRON CORPORATION** [US/US]; 4560 Horton Street, Emeryville, CA 94608-2916 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **GRANDI, Guido** [IT/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **TELFORD, John** [IT/US]; C/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **BENSI, Giuliano** [IT/US]; C/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(74) Agents: **HALE, Rebecca, M.** et al.; Intellectual Property-R-338, Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
3 November 2005
Date of publication of the revised international search report:
15 December 2005

(15) Information about Correction:
see PCT Gazette No. 50/2005 of 15 December 2005, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMMUNOGENIC COMPOSITIONS FOR STREPTOCOCCUS PYOGENES

(57) Abstract: The invention includes a GAS antigen, GAS 40, which is particularly suitable for use either alone or in combinations with additional GAS antigens, such as GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057.

WO 2005/032582 A3

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/024868

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 A61K39/09

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, Sequence Search, WPI Data, PAJ, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE Geneseq [Online] 2 July 2002 (2002-07-02), "Streptococcus polypeptide SEQ ID NO 9188." XP002334315 retrieved from EBI accession no. GSN:ABP30006 Database accession no. ABP30006 the whole document	24-26
A	-& DATABASE EPO Proteins [Online] 2 February 2004 (2004-02-02), "Sequence 7466 from Patent WO0234771." XP002334316 retrieved from EBI accession no. EPOP:CQ650509 Database accession no. CQ650509 the whole document	3,10-17
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 July 2005	28. 10. 05

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Perez, C

4

Form PCT/ISA/210 (second sheet) (January 2004)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	-& WO 02/34771 A (CHIRON S.P.A; THE INSTITUTE FOR GENOMIC RESEARCH; TELFORD, JOHN; MASIG) 2 May 2002 (2002-05-02) claims 1-28 page 2, line 1 - page 10, line 16	1-17, 21-27
X	DATABASE Geneseq [Online] 19 June 2003 (2003-06-19), "Protein encoded by Prokaryotic essential gene #31978." XP002334512 retrieved from EBI accession no. GSN:ABU46451 Database accession no. ABU46451 the whole document	24-26
X	-& WO 02/077183 A (ELITRA PHARMACEUTICALS, INC; WANG, LIANGSU; ZAMUDIO, CARLOS; MALONE, C) 3 October 2002 (2002-10-03) claim 25	24-26
X	DATABASE Geneseq [Online] 2 December 2004 (2004-12-02), "S. pyogenes hyperimmune system reactive antigen Spy0269." XP002334513 retrieved from EBI accession no. GSN:ADR83896 Database accession no. ADR83896 the whole document	24-26
E	-& WO 2004/078907 A (INTERCELL AG; MEINKE, ANDREAS; NAGY, ESZTER; WINKLER, BIRGIT; GELBMANN) 16 September 2004 (2004-09-16) claim 11	24-26
A	FERRETTI J J ET AL: "Complete genome sequence of an M1 strain of Streptococcus pyogenes" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 98, no. 8, 10 April 2001 (2001-04-10), pages 4658-4663, XP002168716 ISSN: 0027-8424 the whole document	1-27

-/--

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/024868

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	OLIVE C ET AL: "Protection of mice from group A streptococcal infection by intranasal immunisation with a peptide vaccine that contains a conserved M protein B cell epitope and lacks a T cell autoepitope" VACCINE, BUTTERWORTH SCIENTIFIC. GUILDFORD, GB, vol. 20, no. 21-22, 21 June 2002 (2002-06-21), pages 2816-2825, XP004357806 ISSN: 0264-410X the whole document -----	1-27

4

Form PCT/ISA/210 (continuation of second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/024868

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0234771 A	02-05-2002	AU 1412702 A	06-05-2002
		CA 2425303 A1	02-05-2002
		EP 1328543 A2	23-07-2003
		MX PA03003690 A	05-05-2004
WO 02077183 A	03-10-2002	NONE	
WO 2004078907 A	16-09-2004	AU 2004218284 A1	16-09-2004